

بسم الله الرحمن الرحيم

السادة أعضاء اللجنة الأعزاء،

نتقدم إلى هيئتكم الموقرة بترشيح الأستاذة الدكتورة منى دياب «لجائزة مجمع الملك سلمان العالمي للغة العربية»، فئة «الأفراد الذين خدموا اللغة العربية وفروعها»، فرع «حوسبة اللغة العربية وخدمتها بالتقنيات الحديثة»،

الأستاذ الدكتور منى دياب:

- مديرة معهد تقنيات اللغة، كلية علوم الحاسب، جامعة كارنيجي ميلون، الولايات المتحدة الأمريكية.
- رئيس أحد فرق البحث العلمي اللغوي بشركة ميتا (فيسبوك).
- الدكتوراه في اللغويات الحاسوبية ومعالجة اللغات الطبيعية، قسم اللغويات، معهد جامعة ميريلاند لدراسات الحاسوب المتقدمة (UMIACS)، جامعة ميريلاند، كوليدج بارك، الولايات المتحدة الأمريكية

الديباجة

تُعدُّ البروفيسور منى دياب من أبرز الباحثين في مجال معالجة اللغات الطبيعية (NLP)، وتعلّم الآلة التطبيقي، والذكاء الاصطناعي. وتتركز أبحاثها العلمية الحالية في المعالجة الآلية للغة، وخصوصا اللغات منخفضة الموارد اللغوية، والتعددية اللغوية، والدرس اللساني الحاسوبي عبر اللغات، والذكاء الاصطناعي. فعلى مدار 20 عاما من حياتها المهنية، قدمت الدكتورة دياب مساهماتٍ كبيرةً في المعالجة الآلية للغة العربية وتقنياتها وتطبيقاتها؛ فهي رائدة على مستوى البحث العلمي في معالجة القضايا العربية؛ مثل المعالجة النحوية التركيبية والدلالية، تحليل المشاعر في النص العربي، وتحليل لغة وسائل التواصل الاجتماعي، وتبديل الشفرات اللغوية داخل النص الواحد (بين اللغات، أو بين اللهجات داخل اللغة الواحدة).

والبروفيسور منى دياب إحدى الشخصيات البارزة التي ساهمت في التأسيس لمجال المعالجة الآلية للغة العربية الحديثة المعاصرة، في إطار مجال معالجات اللغات الطبيعية / الذكاء الاصطناعي. كما تُعتبر مرجعًا أساسيًا في هذا المجال بأكثر من 150 بحثًا علميًا، وأطروحةً، وتقديرًا فنيًا تغطي معظم فروع المعالجة الآلية للغة العربية، وتبتكر المعايير والمنهجيات البحثية. كما قامت ببناء برمجيات تطبيقية في السياقات الأكاديمية، والصناعية، والمؤسسات الحكومية. وتوظّف هذه البرمجيات في التحليل الآلي الصرفي والتركيبى والمعجمي الدلالي للغة العربية، والترجمة الآلية، وتصنيف ونمذجة خطاب الكراهية في النص المكتوب باللغة العربية.. ولعل أحد أبرز أبحاث البروفيسور منى دياب هو تصميم منهجيات لتحسين مقروئية اللغة العربية آليا؛ عبر التركيز على التحليل، والتشكيل الجزئي الذي يؤدي إلى فك اللبس (تشكيل ما يشكل، آليا).

كما أنها رائدة العمل على معالجة تبديل الشفرات اللغوية (بين اللغات، وبين لهجات اللغة الواحدة)؛ عموما، وفي اللغة العربية خصوصا. ولقد كان لبحوثها كبير الأثر في تطوير نماذج عامة للغات مختلفة؛ على أن تركيزها البحثي انصبّ باستمرار على اللغة العربية، في مقابل النماذج الراسخة للمعالجة الآلية للغات الطبيعية. ولقد شملت أبحاث الدكتورة منى كثيرا من جوانب اللغة العربية؛ مثل بناء الموارد اللغوية، وتطوير نماذج للغة العربية [كالتركيب المعجمية الدلالية لأفعال القرآن الكريم]، واللغة العربية الفصحى الحديثة، والأنواع اللغوية غير الرسمية لوسائل التواصل الاجتماعي، والمعروفة أيضا باسم اللهجات العربية.

Mona T. Diab, PhD

Arabic NLP Relevant Bio Synopsis (Details in attached CV)

Honors and Mentions

- October 2022: Featured on the cover of the premier lay scientific publication for youth in Egypt Shabab AlAhrām: <https://gate.ahram.org.eg/News/3717323.aspx>
- April 2022: Selected as one of the top 30 Global AI Scientists of Arab Descent, MIT Technology Review Arabia, <https://technologyreview.ae/ai-person/>
- March 2021: Featured in the Arabic edition of MIT Technology Review Alkhawarizmeya <https://mailchi.mp/c20e4068fc68/51pu4c3xh6-215141>
- Sept. 2019: Selected as one of top 150 leaders and visionaries in AI nationwide to participate in White House AI Summit in Government, Washington DC, USA
- March 2017: 3 Muslim Women in STEM You Should Know About, Teen Vogue, March 2017, <http://www.teenvogue.com/story/muslim-women-in-stem-to-know>
- October 2012: Arbiter/Referee, One of 6 Arab Expat Scientists to decide on Best Scientific Research in Qatar, Qatar Foundation, Arab Expat Society & Scientific Research Forum, Doha, Qatar.
- November 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, Doha, Qatar.
- September 2011: Upper Administration President's Council, Future of Arabic NLP Technologies, King Abdullah University for Science and Technology (KAUST), Thuwal, Saudi Arabia.
- July 2010: Invited by office of the UAE Crown Prince Court for Strategic Planning as One of 12 global Arabic thinkers globally to discuss and chart out the Future of Arabic Language Development, Abu Dhabi, UAE.

Research Accomplishments Highlights

- *H Index 54, >15K citations (Google Scholar, 06/10/2023)*
- **Publications**
 - >250 publications, tutorials/courses, and posters, 169 of which with a focus on Arabic NLP either directly or as a use case study. **3 best paper awards** on Arabic NLP, the latest of which was on Arabic Hate Speech Detection and Classification
 - *Badr AlKhamissi, Mona Diab. (2022) Meta AI at Arabic Hate Speech 2022: MultiTask Learning with Self-Correction for Hate Speech Classification. Proceedings of OSACT Workshop, LREC 2022. Best Paper Award*
 - *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. Best Poster Award*

- *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). Best Paper Award*

- **Arabic NLP Research Areas**

- Original Arabic resource creation with guidelines establishing standards and direction for the field (details in extended CV)
 - *Word level: Mandiac diacritization datasets, Tharwa, Lexicons, Arabic WordNet, unique dialectal repositories for Arabic variants;*
 - *Syntactic structural representations: Arabic treebank splits;*
 - *Semantics representations: Arabic Propbank;*
 - *Arabic dialectal standard orthography guidelines;*
 - *Dialectal Arabic code switching repository*
 - *Machine Translation reference translation for Arabic code switched data*
 - *Political Ideology stance classification data for Egyptian Arabic*
 - *Readability datasets*
 - *Sentiment Analysis*
 - *(crosslingual) Semantic Textual Similarity*
 - *Named Entity Recognition*
 - *Arabic Multiword Expressions Classification*
 - *Annotation Tools*
- Pioneering modeling using latest ML technologies on:
 - *Arabic tokenization, part of speech tagging, syntactic parsing and chunking, named entity recognition*
 - *Arabic linguistic code switching*
 - *Arabic word sense disambiguation*
 - *Arabic subgroup detection*
 - *Arabic sentiment analysis*
 - *Dialect identification*
 - *Arabic (dialectal) translation*
 - *Arabic partial diacritization/vowelization*
 - *Arabic medical pamphlet readability*
 - *Arabic stance and detection*
 - *Arabic Hate Speech detection and classification*
 - *Arabic Multiword expression detection and classification*

- **Released (Robust) technologies (9 systems/tools):**

- *[WASA] Annotation tool for linguistic code switching token level annotation (2018)*
- *[MANDIAC] Arabic Diacritization Annotation Tool (2016)*
- *[DIRA] Dialectal Arabic Information Retrieval Assistant. (2013)*
- *[AIDA] Automatic Identification of Dialectal Arabic. (2013)*
- *[MADAMIRA] A system for morphological analysis and disambiguation, base phrase chunking and named entity recognition. (2013)*
- *[CODACT] A Tool for Arabic Dialect Orthographic Conventionalization (2011)*

- **[AMIRA]** *Automatic Segmentation/Tokenization, Lemmatization, Diacritization, POS Tagging, Base Phrase Chunking/Shallow Parsing of Arabic Text. Current Version 2.0.(2011)*
- **[COLANN]** *COLABA Annotation Interface v.1.0: An annotation interface for Arabic Dialects. Current Version 1.0. (2010)*
- **[NIDA]** *Arabic Named Entity Recognition. Current Version 2.0 (2010)*

Visibility & Outreach

• Keynote Addresses

- Delivered 32 keynote addresses at top scientific venues (major conferences such as EMNLP, KDD, AMTA, ACLing, IEEE-SNAMS, and workshops such as EACL UNLP, WANLP, NAACL WOH, NeurIPS SustainNLP, NeurIPS Muslims in ML) with 31 of them referencing Arabic NLP.
- 14 of these keynotes were delivered in the Middle East (Saudi Arabia, Egypt, Jordan, Morocco, UAE, Palestine, Qatar)

• Invited Talks

- Delivered 67 invited talks addressing audiences in academia (Stanford, Columbia, University of Washington, Princeton), industry (Amazon, Google, IBM) and government (NIST, DARPA, MITRE, IARPA, NSF), 14 of which were delivered in the Middle East and Pakistan

• Panel Moderator/coordinator

- Devised and organized 26 panel discussions at top venues, 9 of which were on Arabic NLP or Low resource language research involving Arabic

• Tutorials and classes

- Tutorial on Arabic Dialect Processing (codesigned and co-taught with Nizar Habash)
 - First designed and delivered in 2007
 - The tutorial runs for 5 hrs and was attended by 100s of people worldwide in person.
 - It was delivered 7 times from 2007-2014 in Morocco, Cairo, USA, Canada
- An A-Z crash course on Statistical NLP
 - Designed and delivered in 2006 and 2007 mainly to audiences in Cairo and Morocco

• Conference and Workshop Organization

- Overall, 52 such initiatives, 22 of which are directly or closely relevant to Arabic NLP (topically)
- 6 of which took place in the MENA region, Qatar, Morocco, Egypt
 - Co-founder of the 1st ACL Arabic NLP Conference (WANLP 2023)

- Co-Chair, 5th Workshop on Computational Approaches to Linguistic Code switching, NAACL 2021
- Co-Chair, 4th Workshop on Computational Approaches to Linguistic Code switching, LREC 2020
- Co-Chair, 3rd Workshop on Computational Approaches to Linguistic Code switching, ACL 2018
- Area Co-Chair Multilingual Processing, ACL 2017
- PC Co-Chair 2nd Workshop on Arabic Natural Language Processing (WANLP), EACL 2017
- Area Co-Chair MT and Multilingual Processing, EMNLP 2017
- PC Co-Chair MT and Multilingual Processing, EACL 2017
- Co-Chair 2nd Workshop on Computational Approaches to Linguistic Code Switching (CALCS), Proceedings of EMNLP 2016, Austin TX, USA
- Co-Chair Workshop on Machine Translation for Semitic Languages (SeMAT), AMTA 2016, Austin TX, US
- Organizer, Mutliword Expressions Special Event, CICLING 15, Cairo, Egypt.
- Co-Chair, First Workshop on Computational Approaches to Linguistic Code Switching, EMNLP 2014, Doha, Qatar.
- Publicity Co-Chair, EMNLP 2014, Doha, Qatar.
- Area Co-Chair, Multilinguality, ACL 2012
- Chair/Co-Organizer of 2nd NSF Funded Workshop on Linguistic Code Switching, 2011
- Chair/Co-Organizer of NSF Funded Workshop on Linguistic Code Switching, 2010
- Area Chair, Multilinguality, EMNLP 2009
- Co-Chair LREC Workshop on Arabic and Local Languages, LREC 2008
- Chair/Co-Organizer of NSF Funded Workshop on Linguistics Code Switching, 2008
- Organizer of SEMEVAL Task on Arabic Semantic Labeling, SemEval 2007
- Co-Chair ACL Workshop on Computational Approaches to Semitic Languages, ACL 2005

Arabic and Arab Talent Mentorship

- **Thesis (sole) supervision**
 - Graduated 15 PhD students and 3 MPhil, 11 of whom worked directly on Arabic NLP.
 - 4 students moved back to MENA region:
 - **Hanan AIDarmaki**, Assistant Professor, Computer Science Dept., MBZUAI, UAE
 - **Fahad AlGhamdi**, Assistant Professor, Computer Science Dept., AlBaha University, Saudi Arabia
 - **Nada AlMarwani**, Assistant Professor, Computer Science Dept., Taibah University, Saudi Arabia
 - **Sawsan AlQahtani**, Research Scientist at SADAIA, Assistant Professor, Computer Science Dept., Princess Noura University, Riyadh, Saudi Arabia

- **Thesis evaluation and examination**
 - 25 Global PhD theses committees, 9 of which are on Arabic NLP
 - 3 of which are in Arab universities based in the MENA region
- **Researcher mentorship**
 - Dr Diab officially mentored and trained (beyond direct supervision of her PhD and MSc students) over 20 research scientists, research engineers, postdoctoral scientists, AI residents and interns in both academia and industry. 13 of whom worked directly on Arabic NLP.

Arabic NLP Community Initiatives and Service

- [05.2005 – 12.2013] (co) Founder (along with Nizar Habash and Owen Rambow) of Columbia Arabic Dialect Modeling Group, Center for Computational Learning Systems (CCLS), Columbia University, the premier global reference point on Arabic NLP since 2005, garnering over \$30M in research funds over a period of ~8yrs.
- [2009-2010] Elected Secretary for ACL Special Interest Group on Semitic Language Processing (SIGSemitic)
- [2011-2017] Elected President for ACL Special Interest Group on Semitic Language Processing (SIGSemitic)
- [2016-2018] Invited Editor in Chief, TAL (the French Reference for NLP) special issue on NLP for Arabic and related languages
- [2021] IEEE Task Force Chair on Large Scale Democratization of Scientific Content leveraging Machine Translation with a focus on Arabic, Chinese, French, Hindi, Spanish
- [2021] Co-founder of the ACL Special Interest Group on Arabic NLP (SIGARAB)
- [2022-present] Elected Treasurer for ACL Special Interest Group on Arabic NLP (SIGARAB)

Expert Consulting on Arabic Language Technologies

- **In the MENA Region**
 - [11.2015-12.2016] Providing Expert Advice on Computational approaches to Readability Complexity Measures in Modern Standard Arabic for Patients' Pharmaceutical Leaflets. Contact: Professor Hend Alkhalifa, King Saud University, Riyadh, Kingdom of Saudi Arabia.
 - [04.2011] Participant, QCRI Social Networks Roundtable, Qatar Computing Research Institute, Qatar Foundation, Doha, Qatar.
 - [01.2011] Participant, QCRI Arabic Language Technologies Roundtable, *Qatar Computing Research Institute, Qatar Foundation*, Doha, Qatar.
 - [10.2007-06.2008] Providing Expert Advice on Arabic Statistical Natural Language Processing. Contact: Dr. Tarek Abbadly. Cairo Microsoft Research Innovation Center, Cairo, Egypt.
 - [04.2006-12.2007] Providing Expert Advice on Statistical Natural Language Processing. Contact: Aly Fahmy and Nayer Wanas. Egyptian Center of Excellence for Data Mining, Ministry of Technology and Information, Cairo Egypt

- **In the USA**

- [04.2010-12.2012] Arabic Propbank, Providing Expert advice predicate argument structures for Arabic. Contact: Professor Martha Palmer, University of Colorado at Boulder, Boulder CO, USA
- [09.2005-12.2007] Arabic WordNet. Providing Expert Advice on Arabic Lexical Semantics for building an Arabic WordNet Repository. Contact: Dr. Christiane Fellbaum, Princeton University, New Jersey, USA
- [04.2006-12.2007] Providing Expert advice on Arabic Lexical Semantics for Machine Translation, Contact: Professor Jaime Carbonell. Meaningful Machines Inc., New York, USA

Professional Positions

- [08.2023--] **Director of Language Technologies Institute**, Tenured Full Professor, School of Computer Science, Carnegie Mellon University, Pittsburgh PA, USA
- [02.2020-08.2023] **Lead Responsible AI Research Scientist**, Meta, Seattle WA, USA
- [06.2017-05.2023] **Tenured Full Professor**, Computer Science Dept. George Washington University, Washington DC, USA
- [10.2017-01.2020] **Principal Scientist**, AWS, Amazon, Seattle WA, USA
- [02.2013 – 06.2023] **Founding Director** of CARE4LANG GWU NLP Lab, Computer Science Dept. George Washington University, Washington DC, USA
- [01.2013-05.2017] **Tenured Associate Professor**, Computer Science Dept. George Washington University, Washington DC, USA
- [02.2005-12.2012] **Research Scientist & Principal Investigator**, Center for Computational Learning Systems, Columbia University, New York NY, USA
- [02.2005-12.2012] **Adjunct Associate Professor**, Department of Computer Science, Columbia University, New York NY, USA

Education & Academic Training

- [01.2004-01.2005] **Postdoctoral Research Scientist**, Linguistics Department and Natural Language Processing Group, Stanford University, Stanford CA, USA. (Mentor: Daniel Jurafsky, MacArthur Grant Recipient 2002-2003)
- [07.2003-12.2003] **Research Associate**, Center for Spoken Language Understanding, University of Colorado at Boulder, Boulder CO, USA. (Mentor: Daniel Jurafsky, MacArthur Grant Recipient 2002-2003)
- [06.1997-08.2003] **Ph.D. Computational Linguistics**, Linguistics Department, University of Maryland Institute for Advanced Computer Studies (UMIACS), University of Maryland at College Park, College Park MD, USA. (Advisor: Philip Resnik)
- [08.1995-05.1997] **M.Sc. Computer Science** (Major: Machine Learning and Artificial Intelligence), Department of Computer Science, The George Washington University, Washington DC, USA. (Advisor: Peter Bock)
- [01.1988-05.1992] **B.Sc. Computer Science**, Department of Computer Science, The American University in Cairo, Cairo, Egypt
- [09.1986-06.1990] **B.Sc. Tourist Guidance** (Major: Egyptology & Archaeology), Faculty of Tourism and Hotel Management, Helwan University, Cairo, Egypt

Detailed CV

Mona T. Diab, PhD

**Director of Language Technologies Institute
(Tenured) Full Professor**
School of Computer Science
Carnegie Mellon University
Pittsburgh, USA
mdiab@andrew.cmu.edu
<http://monadiab.org>

Lead Research Scientist
Meta
Responsible Ads Experiences
Monetization Org
Seattle, WA 98109
mdiab@meta.com

(Arabic focus/highlight in green below)

- Google Scholar citations >14.8K and H-Index 54 across >250 publications

RESEARCH FOCUS & INTERESTS

Current Research Focus

- **Leading Arabic Language Identification, Arabic Hate Speech Modeling and Labeling within Meta**
- **Leading IEEE task force on machine translation of scientific content (Democratization of Science Initiative)**
- **Leading and Driving research as Lead Responsible AI Research Scientist in Responsible AI across Meta, Responsible AI Group, with a special focus on Trustworthiness & Regulatory Compliance (through Dec. 2022)**
- **Leading and Driving research for Responsible NLP in Facebook AI Research (FAIR), with a special focus on Large pretrained multilingual Language Models and Natural Language Generation (through Jan 31st, 2022)**
- **Led Conversational AI within Amazon AWS AI (Project Lex).**

General Research Interests

- **Trustworthy NLP and Responsible AI, Controllable Natural Language Generation, Conversational AI, Generative AI, Cross lingual/multilingual processing, Applied machine learning techniques, Computational lexical semantics/sociolinguistics/pragmatics, Social media/health analytics, Mental Health modeling, low resource language processing, resource building, text analytics, information extraction, sentiment and emotion analysis, Arabic computational linguistics, Democratization of Scientific Content.**

HONORS & MEDIA MENTIONS

1. **October 2022:** Featured on the cover of the premier lay scientific publication for youth in Egypt **Shabab AlAhram**: <https://gate.ahram.org.eg/News/3717323.aspx>
2. **April 2022:** Selected as one of the top 30 Global AI Scientists of Arab Descent, **MIT Technology Review Arabia**, <https://technologyreview.ae/ai-person/>

3. March 2021: Featured in the Arabic edition of **MIT Technology Review Alkhawarizmeya** <https://mailchi.mp/c20e4068fc68/51pu4c3xh6-215141>
4. Sept. 2019: Selected as one of top 150 leaders and visionaries in AI nationwide to participate in **White House AI Summit in Government**, Washington DC, USA
5. May 2017: Behind Every Strong Woman Is...Another Strong Woman: Ten women give thanks to the women who supported them on the way up. **Elle**, May 2017. <http://www.elle.com/life-love/g29746/real-secret-behind-successful-women/?slide=6>
6. March 2017: 3 Muslim Women in STEM You Should Know About, **Teen Vogue**, March 2017, <http://www.teenvogue.com/story/muslim-women-in-stem-to-know>
7. Feb 2008: Diab, Mona *It's About Time*. **CMIC Newsletter**, Volume 1, issue #2. Feb 15th. <http://www.microsoft.com/middleeast/Egypt/CMIC/Newsletter.aspx>

PROFESSIONAL EXPERIENCE

Carnegie Mellon University

06.2023-onwards **Director of Language Technologies Institute, Technical (Tenured) Full Professor**
School of Computer Science
Pittsburgh, PA, USA

Meta

01.2023-present **Lead Research Scientist, Technical Lead**, Responsible Ads Experiences, Responsibility & Privacy, Meta, Seattle WA, USA
 01.2022-12.2022 **Lead Responsible AI Research Scientist, Technical Lead**, Responsible AI Org, Meta, Seattle WA, USA
 07.2021-12.2021 **Lead Research Scientist, Technical Lead LLM**, FAIR ACCEL, Meta, Seattle WA, USA
 02.2020-06.2021 **Lead Research Scientist, Technical Lead, Machine Translation Evaluation, Responsible NLP**, FAIAR, Facebook AI/Meta, Seattle WA, USA

George Washington University

06.2017-05.2023 **(Tenured) Full Professor**, Department of Computer Science, The George Washington University, Washington DC, USA
 01.2013-06.2023 **Director**, GW NLP Lab (CARE4Lang), The George Washington University, Washington DC, USA (~20 active members)
 06.2005-present **Co-Director**, Computational Approaches for Arabic Dialect Modeling (CADIM) Group, Columbia University, The George Washington University, NYU-Abu Dhabi (~10 active members)

Amazon

11.2017-01.2020 **Principal Applied Scientist**, Amazon Web Services (AWS), AI Deep Learning Group for Human Language Technologies, Seattle WA, USA.

01.2013-05.2017 (**Tenured**) **Associate Professor**, Department of Computer Science, The George Washington University, Washington DC, USA

Columbia University

09.2009-12.2012 **Research Scientist (Principal Investigator)**, Center for Computational Learning Systems (CCLS), Columbia University, New York NY, USA

09.2009-12.2012 **Adjunct Associate Professor**, Department of Computer Science, Columbia University, New York NY, USA

09.2007-08.2009 **Adjunct Assistant Professor**, Department of Computer Science, Columbia University, New York NY, USA

02.2005-08.2009 **Associate Research Scientist (Principal Investigator)**, Center for Computational Learning Systems (CCLS), Columbia University, New York NY, USA

EDUCATION & ACADEMIC TRAINING

01.2004-01.2005 **Postdoctoral Research Scientist**, Linguistics Department and Natural Language Processing Group, Stanford University, Stanford CA, USA. (Mentor: Daniel Jurafsky, **MacArthur Grant Recipient 2002-2003**)

07.2003-12.2003 **Research Associate**, Center for Spoken Language Understanding, University of Colorado at Boulder, Boulder CO, USA. (Mentor: Daniel Jurafsky, **MacArthur Grant Recipient 2002-2003**)

06.1997-08.2003 **Ph.D.** Computational Linguistics, Linguistics Department, University of Maryland Institute for Advanced Computer Studies (UMIACS), University of Maryland at College Park, College Park MD, USA. (Advisor: Philip Resnik)

08.1995-05.1997 **M.Sc.** Computer Science (Major: Machine Learning and Artificial Intelligence), Department of Computer Science, The George Washington University, Washington DC, USA. (Advisor: Peter Bock).

01.1988-05.1992 **B.Sc.** Computer Science, Department of Computer Science, The American University in Cairo, Cairo, Egypt.

09.1986-06.1990 **B.Sc.** Tourist Guidance (Major: Egyptology & Archaeology), Faculty of Tourism and Hotel Management, Helwan University, Cairo, Egypt.

ELECTED POSITIONS

01.2023 – present **President**, Association for Computational Linguistics, Special Interest Group for Data (SIG-DAT), EMNLP board

01.2022 – present **Treasurer**, Association for Computational Linguistics, Special Interest Group for Arabic (SIG-Arab), WANLP Conference board

01.2022 – 12.2022 **VP**, Association for Computational Linguistics, Special Interest Group for Data (SIG-DAT), EMNLP board

01.2021 – 12.2021 **VP-Elect**, Association for Computational Linguistics, Special Interest Group for Data (SIG-DAT), EMNLP board

11.2011-12.2017 **President**, Association for Computational Linguistics, Special Interest Group for Semitic Language Processing (SIG-Semitic)

10.2013-09.2016 **President**, Association for Computational Linguistics, Special Interest Group on the Lexicon (SIGLEX)

03.2014-12.2014 **Senator**, University Faculty Senate, GW

01.2012-06.2014 **Executive Board Member**, North American Association for Computational Linguistics (NAACL)

06.2010-09.2013	Secretary , Association for Computational Linguistics Special Interest Group on the Lexicon (SIGLEX)
08.2009-12.2012	Senator , University Senate, one of six senators representing >3000 Research Officers, Columbia University
06.2009-10.2010	Secretary , Association for Computational Linguistics Special Interest Group for Semitic Language Processing (SIG-Semitic)
10.2007-07.2010	Executive Board Member , Association for Computational Linguistics Special Interest Group on the Lexicon (SIGLEX)

Editorial BOARD MEMBERSHIP

04.2021-03.2022	Action Editor, <u>ACL Rolling Review (ARR)</u>
08.2020-08.2021	ACL Committee on Ethical NLP
07.2017-07.2019	Action Editor, <u>Transactions of Computational Linguistics (TACL)</u>
01.2017-12.2019	Board Member, <u>Journal of Natural Language Engineering</u>
10.2016-10.2018	Invited Editor in Chief , <u>TAL (the French Reference for NLP) special issue on NLP for Arabic and related languages</u>
04.2016-12.2019	Board Member, <u>Journal of Computer Speech and Language</u>
07.2013-12.2019	Board Member, <u>Journal of Computational Linguistics</u>
10.2011-12.2019	Board Member, <u>Journal of Language Resources and Evaluation</u>
07.2011-12.2017	Co-Chair, <u>NLP Briefs</u> , Springer Publishers

ADVISORY BOARD MEMBERSHIP

04.2023-present	International Academic Advisory Board, University of British Columbia Language Sciences Institute, Vancouver, BC, Canada
01.2023-present	Advisory Board Member, Center for Human Inspired AI (CHIA), Cambridge University, UK
01.2021-present	ACL Committee on Green NLP
01.2020-present	ACL Committee on Arabic NLP
09.2011-09.2013	Qordoba Inc., Dubai, UAE
09.2011-09.2012	(ISO) Semantic Annotation Framework –Named Entities
05.2010-09.2011	Cairo Microsoft Research Innovation Center (CMIC), Microsoft Research, Cairo, Egypt

PROFESSIONAL CONSULTING

11.2015-12.2016	Providing Expert Advice on Computational approaches to Readability Complexity Measures in Modern Standard Arabic for Patients' Pharmaceutical Leaflets. <i>Contact: Professor Hend Alkhalifa, King Saud University</i> , Riyadh, Kingdom of Saudi Arabia.
04.2010-12.2012	Arabic Propbank, Expert advice predicate argument structures for Arabic. <i>Contact: Professor Martha Palmer, University of Colorado at Boulder</i> , Boulder CO, USA
10.2007-06.2008	Providing Expert Advice on Arabic Statistical Natural Language Processing. <i>Contact: Dr. Tarek Abbady. Cairo Microsoft Research Innovation Center</i> , Cairo, Egypt.
04.2006-12.2007	Providing Expert Advice on Statistical Natural Language Processing. <i>Contact: Aly Fahmy and Nayer Wanas. Egyptian Center of Excellence for Data Mining, Ministry of Technology and Information</i> , Cairo Egypt

- 09.2005-12.2007 Providing Expert Advice on Arabic Lexical Semantics for building an Arabic WordNet Repository. *Contact: Dr. Christiane Fellbaum, Princeton University, New Jersey, USA*
- 04.2006-12.2007 Providing Expert advice on Arabic Lexical Semantics for Machine Translation, *Contact: Professor Jaime Carbonnell. Meaningful Machines Inc., New York, USA*

ACADEMIC/SCIENTIFIC INITIATIVES

[Total: 8]

1. [09.2021] Founder and Initiator of **ACL 60-60 Initiative** trailblazing the translation of speech and text modalities of scientific content into 60 languages within a year 2022-2023
<https://www.2022.aclweb.org/dispecialinitiative>
2. [11.2014] **Initiator of SEAS Big Data Hub proposal to NSF** (with Michael Plesniak), setting up GW as a regional force in Big Data Sciences with an eye on technology transfer. Currently, GW is part of the SW Regional BD hub.
3. [05.2014] **Co-initiator Data Science Track** (with Evan Drumwright, Claire Monteleoni, Nan Zhang, Abdou Youssef). Department of Computer Science, The George Washington University, Washington DC, USA.
4. [07.2011] **Founding member, *SEM Conference** (with Eneko Agirre, Johan Bos). A unified consolidated top tier computational semantics scientific venue currently in 12th edition.
5. [04.2008] Establishing **Computational Approaches of Linguistic Code Switching** as a track in NLP culminating in a workshop with shared tasks from 2014 onwards (currently in 5th edition (jointly with Thamar Solorio, Julia Hirschberg, Pascale Fung)
6. [07.2007] **Founding member, Columbia Arabic Language Institute for Modeling and Technology (CALIMAT)** (with Hatim Diab, Nizar Habash, Owen Rambow, David Waltz). Seeking joint institutional collaborations with the Middle East and North African countries. Center for Computational Learning Systems (CCLS), Columbia University, New York, NY, USA.
7. [05.2007] Establishing **Semantic Textual Similarity (STS)** as a new area of research in Computational Linguistics (jointly with Eneko Agirre)
8. [06.2005] **Founding member, Columbia Arabic Dialect Modeling Group, currently Computational Approaches for Arabic Dialect Modeling (CADIM)** (with Nizar Habash and Owen Rambow), Center for Computational Learning Systems (CCLS), Columbia University, New York, NY, USA.

KEYNOTE ADDRESSES

[Total: 32]

1. [05.2023] **2nd Ukrainian NLP, EACL 2023**, Croatia
2. [12.2022] **EMNLP 2022**, Abu Dhabi, UAE
3. [07.2022] **WOAH Workshop, NAACL 2022**, Seattle, WA, USA
4. [07.2022] **DataPerf Workshop, ICML 2022**, Maryland, MD, USA
5. [11.2021] **Conversational AI Workshop, EMNLP 2021**, Punta Cana, Dominican Republic
6. [11.2021] NLP Track, **Open Data Science Conference (ODSC) West 2021**, Miami Florida, USA
7. [08.2021] **Low Resource Machine Translation (LoResMT) Workshop, MT Summit 2021**, Virtual
8. [07.2021] **International Conference on Information Technology (ICIT)**, Jordan

9. [06.2021] **5th International Conference on AI in Computational Linguistics (ACLing) 2021**, virtual
10. [12.2020] **International Conference on Islamic Applications in Computer Science and Technologies (IMAN)**, virtual
11. [12.2020] **IEEE International Conference on Social Networks Analysis, Management and Security (SNAMS-2020)**, Paris, France
12. [11.2020] **Machine Learning in Arabic, Women to Impact Workshop (WTI)**, KAUST, Saudi Arabia
13. [10.2020] **The 14th biennial conference of the Association for Machine Translation in the Americas (AMTA 2020)**, virtual
14. [10.2019] **The 7th International Conference on Arabic Language Processing (ICALP2019)**, Nancy, France.
15. [10.2019] **The Second International Conference on Trends in Computing Sciences (ICTCS'19)**, Amman, Jordan
16. [07.2019] **WiDening NLP ACL Workshop (WiNLP)**, Florence, Italy.
17. [06.2019] **The 4th Swiss Text Analytics Conference, SwissText 2019**, Winterthur, Switzerland.
18. [05.2018] **3rd Workshop on Open-Source Arabic Corpora and Processing Tools (OSACT3)**, LREC 2018, Miyazaki, Japan.
19. [04.2017] **Trends in Arabic Pedagogy, 3rd Roundtable Workshop**, Georgetown University in Qatar, Qatar.
20. [12.2016] **COLING 2016 Third Workshop on NLP for Similar Languages, Varieties and Dialects (VARDIAL 2016)**, Osaka, Japan.
21. [04.2015] **16th International Conference on Intelligent Text Processing and Computational Linguistics (CICLING 15)**, Cairo, Egypt.
22. [04.2015] **1st Arabic Computational Linguistics Conference (ARCLing15)**, Cairo, Egypt.
23. [12.2014] **9th IEEE International Conference on Computer Engineering and Systems (ICCES 14)**, Rabat, Morocco.
24. [03.2014] **3rd Palestinian Symposium on Computational Linguistics and Arabic Content**, Birzayt University, West Bank, Palestine.
25. [11.2013] **Joint Symposium on Semantic Processing, organized by the European Union funded projects EXCITEMENT and B-CROCE hosted by FBKI**, Trento, Italy.
26. [03.2013] **27th Arabic Linguistic Symposium (ALS 13)**, Indiana IN, USA.
27. [12.2011] **11th International Language Engineering Conference**, Cairo, Egypt.
28. [09.2011] **IEEE 5th International Conference on Semantic Computing (ICSC)**, Palo Alto CA, USA.
29. [04.2007] **IEEE Information and Communication Technologies International Symposium (ICTIS07)**, Fez, Morocco.
30. [11.2006] **6th Conference on Language Engineering**, Cairo, Egypt
31. [04.2006] **EACL Workshop on Cross Language Knowledge Induction**, Trento, Italy
32. [03.2006] **Arabic Linguistics Society (ALS) Conference**, Kalamazoo MI, USA

INVITED TALKS

[Total: 67]

1. [5.2023] **Webinar to C-Level Executives, LyRise**, Egypt & USA
2. [5.2023] **University of Edinburgh**, UK
3. [3.2023] **Cambridge University**, UK
4. [3.2023] **AI2**, Seattle, WA, USA
5. [3.2023] **University of Washington**, Seattle, WA USA

6. [1.2023] **Spotify**, NYC, USA
7. [11.2022] Distinguished Lecture Series, CRFM, Human AI Institute, **Stanford University**, CA USA
8. [11.2022] NLP Seminar Series, **Stanford University**, CA USA
9. [10.2022] Distinguished Lecture Series, Cognitive Science Department, **Rensselaer Polytechnique Institute**, NY, USA
10. [11.2021] RAISE Program, Information School, **University of Washington**, Seattle, USA (Host: Prof. Chirag Shah)
11. [03.2021] MIDAS Distinguished Lecture Series, **University of Michigan**, USA (Host: Prof. Rada Mihalcea)
12. [03.2021] Data Science Institute, **Columbia University**, USA (Host: Prof. Smaranda Muresan)
13. [03.2021] LangSci Distinguished Lecture Series, **University of British Columbia**, Canada (Host: Prof. Muhammad AbdulMageed)
14. [02.2021] NLP Lecture Series, **Imperial College**, London, UK (Host: Prof. Lucia Specia)
15. [12.2020] Muslims in ML Workshop, **NeurIPS 2020**
16. [11.2020] SustainNLP Workshop, **EMNLP 2020**
17. [10.2020] UIUC NLP Seminar Series, **University of Illinois University of Chicago** (Host: Prof. Heng Ji)
18. [10.2020] STEM:NEXT WAVE2 - STEM Superstar #6, **National STEM School**, Pakistan
19. [09.2020] Muslim Women in Science & Technology, **Khwarizmi Science Society**, Pakistan
20. [08.2020] Applied Data Science Talks, **KDD 2020**
21. [07.2020] University of Houston, Texas, **NSF REU Program**, Houston, TX, USA
22. [04.2017] **Amazon INC**, Seattle, WA, USA
23. [01.2017] Loudoun County, **Virginia Economic Development**, GW Virginia Campus, VA, USA
24. [12.2016] Department of Computer Science and Linguistics, **Georgetown University**, Washington DC, USA.
25. [09.2016] Department of Computer Science, **CLIP Lab, UMIACS, University of Maryland College Park**, Maryland MD, USA.
26. [09.2016] Department of Computer Science, **Johns Hopkins University**, Baltimore MD, USA.
27. [04.2016] Department of Computer Science, **King Saud University**, Riyadh, Kingdom of Saudi Arabia.
28. [04.2016] Department of Computer Science, **AI Imam University**, Riyadh, Kingdom of Saudi Arabia.
29. [04.2016] Languages & Linguistics Department, **Princess Noura University**, Riyadh, Kingdom of Saudi Arabia.
30. [03.2016] School of Engineering and Applied Science, **King Abdullah University for Science & Technology**, Thuwal, Kingdom of Saudi Arabia.
31. [05.2015] **IBM Watson and Cognitive Computing Institute**, New York NY.
32. [03.2013] **MITRE Corporation**, Tyson's Corner VA, USA.
33. [03.2013] Computer Science Symposium, **Simon Fraser University**, British Columbia, Vancouver BC, Canada.
34. [02.2013] Computer Science Symposium, **CLIP Lab, UMIACS, University of Maryland College Park**, College Park MD, USA.
35. [09.2011] Upper Administration, **King Abdullah University for Science and Technology (KAUST)**, Thuwal, Saudi Arabia.
36. [09.2011] Mathematical & Computer Sciences School, **King Abdullah University for Science and Technology (KAUST)**, Thuwal, Saudi Arabia.

37. [10.2011] **NSF-sponsored Workshop on Collaborative Resource Development and Delivery**. Organized by Nancy Ide, Vassar College, New York NY, USA.
38. [04.2011] **Distinguished Lecture Symposium Series, University of Washington**, Seattle WA, USA.
39. [04.2011] **Microsoft Research**, Redmond WA, USA.
40. [12.2010] Computer Science Department, **University of Alabama**, Birmingham AL, USA.
41. [07.2010] **Workshop on the Future of Arabic Language Development**, organized by the **UAE Crown Prince Court for Strategic Planning**, Abu Dhabi, UAE.
42. [06.2010] **IBM TJ Watson**, Hawthorne NY, USA.
43. [05.2010] School of Computing and Mathematical Sciences, **King Abdullah University for Science and Technology, KAUST**, Thuwal, Saudi Arabia.
44. [04.2010] Computer Science Department, **University of Trento**, Trento, Italy.
45. [06.2009] **NAACL Workshop on Semantic Evaluations (SEMEVAL 09)**, Boulder CO, USA.
46. [05.2009] Departamento de Sistemas Informaticos y Computacion, **Universidad Politecnica de Valencia**, Valencia, Spain.
47. [04.2009] **IRST & Computer Science Department, University of Trento**, Trento, Italy.
48. [08.2008] Department of Informatics, **University of Sussex**, Brighton, England.
49. [05.2008] Computer Science Department, **Faculty of Science, University of Mohamed V**, Rabat, Morocco.
50. [05.2008] Departamento de Sistemas Informaticos y Computacion, **Universidad Politecnica de Valencia**, Valencia, Spain.
51. [04.2008] Linguistics Department, **Georgetown University**, Washington DC, USA.
52. [12.2006] **Cairo Microsoft Innovation Center (CMIC), MSR**, Cairo, Egypt.
53. [09.2006] Computer Science Department, **University of North Texas**, Dallas TX, USA.
54. [12.2005] Linguistic Data Consortium, **University of Pennsylvania**, Philadelphia PA, USA.
55. [10.2005] Computer Science Department, **New York University**, New York NY, USA.
56. [03.2005] Computer Science Department, **Columbia University**, New York NY, USA.
57. [02.2005] Linguistics Department, **Ohio State University**, Columbus OH, USA.
58. [02.2005] Linguistics Department, **University of Texas Austin**, Austin TX, USA.
59. [01.2005] **Yahoo Research Labs, Overture**, Pasadena CA, USA.
60. [12.2004] **Language Computer Corporation (LCC)**, Dallas TX, USA.
61. [12.2004] Natural Language Group, **Microsoft Research (MSR)**, Redmond WA, USA.
62. [08.2004] Department of Informatics, **University of Edinburgh**, Edinburgh, Scotland.
63. [05.2004] Linguistic Data Consortium, **University of Pennsylvania**, Philadelphia PA, USA.
64. [05.2004] **IBM T. J. Watson Research Center**, Hawthorne NY, USA.
65. [11.2003] Linguistics Dept. & Institute for Cognitive Science, **University of Colorado, Boulder**, Boulder CO, USA.
66. [11.2002] Laboratoire Perception, Systèmes, Information (PSI), **Université de Rouen**, Rouen, France.
67. [11.2002] NLP Lab, Information Sciences Institute (ISI), **University of Southern California**, Marina Del Rey CA, USA.

INVITED PANELIST/MODERATER & COORDINATOR

[Total: 26]

1. [04.2023] Moderator/Coordinator, Evaluation Measures and Frameworks for Responsible Generative AI, **Responsible AI Leadership: A Global Summit on Generative AI, World Economic Forum**, San Francisco CA, USA

2. [11.2022] Moderator/Coordinator, Fairness in NLP, Conference Special Theme, **AAACL 2022**
3. [10.2022] Invited Panelist, Industry Government Panel, **NSF sponsored workshop on Ethical AI Design**, Virtual.
4. [10.2021] Moderator/Coordinator, Privacy and Ethical AI in Academia and Industry Research, **WeCNLP 2021**
5. [08.2021] Participant, **Green NLP**, ACL2021
6. [06.2021] Participant, [Careers in NLP](#), **Industry Panel**, NAACL 2021
7. [06.2021] Moderator/Coordinator, Machine Translation of Code Switched Data: Evaluation Challenges and Opportunities, **5th workshop on Computational Approaches to Linguistic Code Switching (CALCS)**, NAACL 2021
8. [06.2021] Moderator, Arabic NLP in Industrial Settings, WANLP 2021
9. [05.2021] Participant, Industry vs. Academia, **AiXplain**
10. [12.2020] Participant, Workshop on Muslims in ML, **NeurIPS 2020**
11. [11.2020] Participant, SustainNLP Workshop, **EMNLP 2020**
12. [11.2020] Moderator/Coordinator, **Ethics Panel: Publishing in an era of Responsible AI: How can NLP be proactive? Considerations and Implications**; **EMNLP 2020**
13. [08.2019] Participant, **REL4NLP Workshop**, **ACL 2019**, Florence Italy
14. [11.2016] Participant, **Wardat Workshop**, by invitation only, NYU-AD, Abu Dhabi, UAE
15. [05.2012] Participant, **Computational lexical resources and interoperability**, *Language Resources and Evaluation Conference (LREC)*, Istanbul, Turkey.
16. [10.2012] Arbiter/Referee, **One of 6 Arab Expat Scientists to decide on Best Scientific Research in Qatar**, *Qatar Foundation, Arab Expat Society & Scientific Research Forum*, Doha, Qatar.
17. [11.2011] Participant, **One of 50 Arab Expat Scientists to discuss and advise on Future of Scientific Research Strategy in Qatar**, *Qatar Foundation, Arab Expat Society & Scientific Research Forum*, Doha, Qatar.
18. [10.2011] Participant, **Future of Semantics in Machine Translation**, *DARPA BOLT PI Kick-off Meeting*, Tampa FL, USA.
19. [09.2011] Participant, **Adjectives Workshop**, NSF, Organized by Christiane Fellbaum. (Princeton University), Arlington VA, USA.
20. [07.2011] Participant, **EMNLP Workshop on Geometrical Methods in Semantics (GEMS)**. *Empirical Methods For Natural Language Processing (EMNLP 11)*, Edinburgh, UK.
21. [04.2011] Participant, **NSF-funded SILT Workshop Interoperable Semantic Annotation**. Organized by Nancy Ide. (Vassar College) and James Pustejovsky. (Brandeis University), Waltham MA, USA.
22. [04.2011] Participant, **QCRI Social Networks Roundtable**, Qatar Computing Research Institute, Qatar Foundation, Doha, Qatar.
23. [01.2011] Participant, **QCRI Arabic Language Technologies Roundtable**, *Qatar Computing Research Institute, Qatar Foundation*, Doha, Qatar.
24. [08.2010] Participant, **COLING Workshop on Multiword Expressions (MWE 10)**, Beijing, China.
25. [07.2010] Participant, **ACL Workshop on Semantic Evaluations (SEMEVAL)**, Upsala, Sweden.
26. [07.2010] Participant, **Workshop on the Future of Arabic Language Development**, *organized by the UAE Crown Prince Court for Strategic Planning*, Abu Dhabi, UAE.

RESEARCH EXPERIENCE

PAST RESEARCH GRANTS

[Total: 25 grant activities amounting to \$28.7M, Diab's portion \$10M]

1. **[DARPA-LORELEI]** *Sentiment and Emotion Detection in Low Resource Languages (SEEM)* (GW-PI; subaward from Columbia University Julia Hirschberg & Kathy Mckeown (PI)), Total: \$1.75M, GW portion \$236,744K, Jul 2015 – Dec. 2019.
2. **[DARPA-DEFT]** *DRATS: Detecting Relations and Anomalies in Text and Speech*, (GW-PI; subaward from Columbia University, Columbia PI: Owen Rambow) Total \$5.5M+\$350K supplement (supplement from May 2017-Nov 2017), Diab's portion: \$934,107+\$94,433K supplement, GW portion \$707,384K+\$94,433K supplement, Aug 2013 – Nov 2017.
3. **[NSF CRI-P]** *A Unified Universal Multiword Expression Repository*, (PI), Total: \$100K, GW portion \$100K, Jun 2015 –Aug 2017.
4. **[NSF-Large CRI]** *Creating Resources for Multilingual Linguistic Code Switching*, (PI; with co-PI Julia Hirschberg from Columbia University, collaborative with Co-PI Tamar Solorio, U of Houston, GW Prime institution), Total: \$785K, GW portion \$400K, Feb 2013 –May 2017.
5. **[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 \$1.05M, GW portion \$361,771K, May 2014 – Mar 29, 2017.
6. **[Google Faculty Research Award]** *Tharwa++: Building a multidialectal Arabic Lexical Repository*, (PI), Total: \$65K+\$10K in Google credits, GW portion \$75K, Sep 2015 –Dec 2016.
7. **[Google Faculty Research Award]** *Nuanced Sentiment and Perspective Analysis for Arabic Social Media Text*, (PI), GW portion \$71,331K, Dec 2014 –Dec 2015
8. **[DARPA-BOLT]** *BBN-Astral* (PI; collaborative with Columbia University Nizar Habash and Owen Rambow), Total \$5.5M, Diab's portion: \$1.833M, GW portion \$241,570, Oct 2011 –Mar 2014
9. **[DARPA-BOLT]** Subcontract to GW from BBN – *Dialectal Tool and Resource Creation*, (PI; collaborative with Columbia University Nizar Habash and Owen Rambow), Total \$1.05M, Diab's portion: \$350K, GW portion \$301,313.00, Feb 2014 –May 2015
10. **[TSWG-DOD/ODNI-NADIA]** *Natural Language Processing for Dialectal Arabic-Industrialization and Insertion*, (PI; with Owen Rambow and Nizar Habash), Total \$898K, Diab's portion: \$299.33K, Apr 2012 –Jun 2013
11. **[DARPA-MADCAT]** Subcontract from BBN–*Multilingual Automatic Document Classification Analysis and Translation for Arabic OCR*, (Co-PI; with Nizar Habash (PI)) Total \$750K, Diab's portion: \$375K, Apr 2008 –Aug 2013
12. **[DARPA-BOLT]** Subcontract to SRI–*Semantic Machine Translation for Arabic Dialects*, (CCLS-PI; collaborative with Kathy Mckeown, Michael Collins, Julia Hirschberg), Total \$1.7M, Diab's portion: \$425K, Nov 2011 –Aug 2013
13. **[DARPA-BOLT]** Subcontract to LDC–*Resource Guideline Creation for Arabic Dialects*, (Co-PI; with Nizar Habash (PI) and Owen Rambow), Total \$500K, Diab's Portion: \$166.67K Nov 2011 –Dec 2012
14. **[DARPA-GALE]** Subcontract to IBM–*Pilot Unified Framework for Semantic Textual Similarity*, (PI), Total \$25K, Sep 2011 –Apr 2012
15. **[NSA-TTO8]** Subcontract to JHU-COE–*Power and Modality Analysis and Prediction in Social Communication*, (Co-PI; with Owen Rambow (PI)) Total \$345K, Diab's Portion: \$172,500, Sep 2009 –Dec 2012
16. **[IARPA-SCIL]** *Power Rifts and Shifts in Social Communication*. (Co-PI; with Kathy McKeown and Owen Rambow (PI)) Total \$3.210M, Diab's portion: \$1.07M, Aug 2009 –Oct 2012
17. **[NSF-CRI Planning]** *Linguistic Code Switching Pilot Annotation*. (PI; with Julia Hirschberg)

- Total \$78K, Diab's portion: \$55K, Mar 2010–Mar 2011
18. [TSWG-DOD/ODNI-COLABA] Subcontract to ACXION–*Cross language blog alerts–Arabic Blog Processing*. (PI; with Nizar Habash and Owen Rambow) Total \$820K, Diab's portion: \$310K, (CAT/CAIM support), Aug 2009–Apr 2011
 19. [DARPA-GALE] Subcontract to IBM-Rosetta–*Arabic to English Machine Translation*. (Co-PI; with Owen Rambow (PI) and Nizar Habash) Total \$1.170M, Diab's portion: \$390K, May 2008–Apr 2011
 20. [NSA-TTO8] Subcontract to JHU-COE–*Committed Belief and Modality in Multigenres Text*. (Co-PI; with Owen Rambow (PI)) Total \$36K, Diab's portion: \$18K, Aug 2007 –Aug 2010
 21. [DARPA-GALE] Subcontract to SRI-Nightingale–*Arabic to English Machine Translation*. (Co-PI; with Owen Rambow (PI) and Nizar Habash) Total \$1.5M, Diab's portion: \$500K, Sep 2005–Apr 2008
 22. [NSF-SGER] *Automatic Processing of Natural Language Code-Switching*. (PI; with Owen Rambow) Total \$41K, Diab's portion: \$30K, Sep 2007–Aug 2008
 23. [NSF-SGER] *Inducing Multiword Expressions from Comparable Corpora*. (PI) Total \$19K, Sep 2006–Feb 2007
 24. [NSF-JHU Summer Workshop] *Arabic Dialect Parsing*. (PI) Total \$15K, Jun 2005–Jul 2005
 25. [DARPA-GALE] Subcontract to University of Colorado–OntoNotes. *Building an Arabic PropBank*. (PI) Total \$15K, Sep 2005–Feb 2006

TEACHING EXPERIENCE

Average teaching score across all classes at GW from Spring 2013- Spring 2017 is 4.77/5.0.

FORMAL SEMESTER LONG CLASSES

[Total: 10 classes/5 courses]

1. [Spring.2017] **CSCI6709/3709: DeepQA IBM Watson Technology**, Department of Computer Science, GW
2. [Fall.2016] **CSCI6709/3709: Introduction to Statistical Natural Language Processing**, Department of Computer Science, GW
3. [Spring.2014] **CSCI6709/3709: Introduction to Statistical Natural Language Processing**, Department of Computer Science, GW
4. [Fall.2013] **CSCI1112: Data Structures and Algorithms I in Java**, Department of Computer Science, GW
5. [Spring.2013] **CSCI 6709/3709: Introduction to Statistical Natural Language Processing**, Department of Computer Science, GW
6. [Fall.2009] **Arabic Computational Linguistics**, Linguistics and Computer Science Departments, Georgetown University.
7. [Spring.1997] **C as a 2nd Language**, Department of Computer Science, GW
8. [Fall.1996] **C as a 2nd Language**, Department of Computer Science, GW
9. [Spring.1996] **C as a 2nd Language**, Department of Computer Science, GW
10. [Fall.1995] **C as a 2nd Language**, Department of Computer Science, GW

TUTORIALS

[Total: 10 Tutorials]

1. [10.2014] **Arabic Dialect Processing (5 hrs)**, *co-designed and co-taught with Nizar Habash*, Empirical Methods for Natural Language Processing (EMNLP), Doha, Qatar. Attended by 35 people.
2. [06.2012] **Arabic Dialect Processing (5 hrs)**, *co-designed and co-taught with Nizar Habash*, North American Association for Computational Linguistics (NAACL). Montreal, Canada. Attended by 40 people.
3. [04.2010] **Arabic Dialect Processing (5 hrs)**, *co-designed with Nizar Habash*, Georgetown University Round Table (GURT), Washington DC. Attended by 60 people.
4. [04.2009] **Arabic Dialect Processing (5 hrs)**, *co-designed and co-taught with Nizar Habash*, 2nd International Conference on Arabic language Resources and Tools. (MEDAR), Cairo, Egypt. Attended by over 100 people.
5. [08.2008] **Arabic Dialect Processing (5 hrs)**, *co-designed and co-taught with Nizar Habash*, American Machine Translation Association (AMTA), Boston MA, USA. Attended by over 30 people
6. [05.2008] **Arabic Dialect Processing (5 hrs)**, *co-designed and co-taught with Nizar Habash*, Language Resources & Evaluation Conference (LREC), Marrakech, Morocco. Attended by 45 people.
7. [06.2007] **Arabic Dialect Processing (5 hrs)**, *co-designed and co-taught with Nizar Habash*, Human Language Technologies-North American Association for Computational Linguistics (HLT-NAACL). Rochester NY, USA. Attended by 45 people.
8. [03.2007] **An A-Z Statistical NLP Crash Course (6 hrs)**, IEEE Information and Communication Technologies International Symposium (ICTIS07). Fez, Morocco. Attended by 40 people.
9. [09.2006] **Statistical Natural Language Processing: An A-Z Crash Course**. Faculty of Engineering and Computer Studies, Cairo University, Cairo, Egypt. Attended by over 45 people.
10. [02.2005] **Introduction to Computational Linguistics (2 hrs)**, *Linguistics Department*, University of Texas at Austin, Austin TX, USA. Attended by 100 people.

RESEARCH SUPERVISION

RESEARCH SCIENTIST MENTORSHIP

[Total: 5 RS]

1. [04.2015-06.2019] **Mahmoud Ghoneim, PhD**
Received PhD Jul 2012: Faculty of Engineering, Cairo University, Egypt
Research Focus: Building shallow parsing technology and incorporating semantics in MT
Current Position: NLP Manager, BP, Houston, Texas, USA
2. [11.2015-12.2018] **Abdelati Hawwari, PhD**
Received PhD Jul 2012: Faculty of Literature & Linguistics, Cairo University, Egypt
Research Focus: Building Lexical Resources for MWE and Computational Lexicography.
Current Position: Lead Language Taxonomist, Apple. Bay Area, CA, USA
3. [Spring 2015-07.2018] **Mohammed AlBadrashiny**
Research Topic: Language Identification and Spelling error correction
Current Position: Senior Research Engineer, AIXplain, New York, NYC, NY, USA

4. [02.2014-11-2014] **Mohammed Attia, PhD**
 Received PhD Jul 2008: University of Manchester, UK.
 Research Focus: Building Lexical Resources and Spelling error correction
 Current Position: *Research Scientist, Google Inc., New York City, NY, USA*
5. [Spring 2012-Fall 2015] **Ali Elkahky**
 Research Topic: Unified models for dialectal processing
 Department of Computer Science, Columbia University, New York NY, USA
 Current Position: *Research Scientist, Apple. Seattle, WA USA*

POSTDOCTORAL MENTORSHIP

[Total: 6]

1. [09.2015-07.2019] **Efsun Kayi, PhD**
 Received PhD Jul 2014: Department of Computer Science, George Washington University, Washington DC, USA
 Research Focus: Automatic Detection of Psychological Disorders (Schizophrenia and PTSD) in Social Media.
 Current Position: *Research Scientist, Applied Physics Lab (APL), Johns Hopkins University, Maryland, USA*
2. [07.2012-03.2015] **Mahmoud Ghoneim, PhD**
 Received PhD Jul 2012: Faculty of Engineering, Cairo University, Egypt
 Research Focus: Integrating Semantic Knowledge in Statistical Machine Translation
 Current Position: *Research Scientist Manager, BP, Houston, Texas, USA*
3. [03.2012-10.2015] **Abdelati Hawwari, PhD**
 Received PhD Jul 2012: Faculty of Literature & Linguistics, Cairo University, Egypt
 Research Focus: Building Lexical Resources and Computational Lexicography.
 Current Position: *Lead Language Taxonomist, Apple. Bay Area, CA, USA*
4. [05.2010-04.2011] **Davide Picca, PhD**
 Received PhD Apr 2010: University of Lausanne, Switzerland
 Research Focus: Unsupervised semantic modeling for comparable corpora
 Current Position: *Research Technical Laison, University of Lausanne, Switzerland*
5. [05.2009-04.2011] **Marine Carpuat, PhD**
 Received PhD Apr 2008: HKUST, Hong Kong
 Research Focus: Incorporating Lexical Semantics in Statistical Machine Translation
 Current Position: *Associate Professor, Department of Computer Science, University of Maryland College Park, MD, USA.*
6. [09.2009-09.2010] **Yassine Benajiba, PhD**
 Received PhD May 2009: Polytechnical University of Valencia, Spain
 Research Focus: Shallow syntactic parsing in Arabic and integration in SMT and IE

Current Position: Senior Research Scientist, Amazon AWS AI, New York NY, USA

GRADUATE & UNDERGRADUATE RESEARCH SUPERVISION

1. [01.2017] **Vivek Gupta**
Research Topic: Building an evaluation interface for SEEM Lorelei
Department of Computer Science, MSc, graduated Spring 2017

THESIS SUPERVISION

Graduated

[Total: 18; 15 PhD and 3 MSc]

1. [04.2022] **Pedram Hosseini, PhD**
Thesis Topic: Advancing Identification of Causal Relations in Natural Language
Department of Computer Science, The George Washington University, DC, USA
Joint Supervision with Prof. David Broniatowsky
2. [11.2020] **Shabnam Tafreshi, PhD**
Thesis Topic: Multilingual and Cross Lingual Affect and Emotion Detection
Department of Computer Science, The George Washington University, DC, USA
Current Position: Postdoctoral Scholar, University of Maryland College Park
3. [05.2020] **Nada Almarwani, PhD**
Thesis Topic: Non parameterized Sentence Embeddings
Department of Computer Science, The George Washington University, DC, USA
Current Position: Assistant Professor, Computer Science Dept., Taibah University, Saudi Arabia
4. [11.2019] **Fahad Alghamdi, PhD**
Thesis Topic: Computational Approaches to Linguistic Code Switching
Department of Computer Science, The George Washington University, DC, USA
Current Position: Assistant Professor, Computer Science Department, AlBaha University, Saudi Arabia
5. [11.2019] **Sawsan Alqahtani, PhD**
Thesis Topic: Diacritic Restoration for low resource languages
Department of Computer Science, The George Washington University, DC, USA
Current Position: Research Scientist, SADAIA; Assistant Prof. Princess Noura University, Riyadh, Saudi Arabia
6. [09.2019] **Maryam Aminian, PhD**
Thesis Topic: Cross-Lingual Transfer of Semantics in low resource settings
Department of Computer Science, The George Washington University, DC, USA
Current Position: undecided
7. [06.2019] **Sardar Hamidian, PhD**
Thesis Topic: Automatic identification and Classification of misinformation
Department of Computer Science, The George Washington University, DC, USA
Current Position: Research Scientist, Comcast Inc., Washington DC, USA
8. [03.2019] **Hanan AlDarmaki, PhD**
Thesis Topic: Cross Language Sentence Modeling
Department of Computer Science, The George Washington University, DC, USA
Current Position: Assistant Professor, Computer Science Department, MBZUAI, Abu Dhabi, UAE
9. [12.2017] **Ayah Zirikly, PhD**

- Thesis Topic: Tracking Named Entities in Multilingual and Multigenre Settings*
 Department of Computer Science, The George Washington University, DC, USA
 Current Position: Postdoctoral Research Scientist, NIH, Bethesda, MD, USA
10. [12.2017] **Wenchao Mo, MSc**
Thesis Topic: Developing semantics based machine translation evaluation metrics
 Department of Computer Science, The George Washington University, DC, USA
 11. [12.2017] **Sai Li, MSc**
Thesis Topic: Automatic augmentation of lexical resources leveraging deep learning methods
 Department of Computer Science, The George Washington University, DC, USA
 12. [09.2017] **Hebatullah Elfardy, PhD**
Thesis: Perspective Identification in multilingual and multigenre contexts
 Department of Computer Science, Columbia University, New York NY, USA
 Current Position: Machine learning Research Scientist, Amazon Inc.
 13. [07.2015] **Vinodkumar Prabhakaran, PhD**
 Joint Supervision with Owen Rambow
Thesis: Social Power in Interactions: Computational Analysis and Detection of Power Relations
 Department of Computer Science, Columbia University, New York NY, USA
 Current Position: Research Scientist, Google Inc., Palo Alto, CA, USA
 14. [04.2015] **Weiwei Guo, PhD**
Thesis: 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
 15. [12.2014] **Muhammad Abdul-Mageed, PhD**
 Joint Supervision with Sandra Kuebler
Thesis: Sentiment Analysis in Arabic Social Media Text
 Linguistics Department, Information Sciences Department
 Indiana University, Indiana IN, USA
 Current Position: Associate Professor, University of British Columbia, Canada
 16. [12.2012] **Pradeep Dasigi, MSc**
Thesis: Dialectal Arabic Orthography Normalization
 Department of Computer Science, Columbia University, New York NY, USA
 Current Position: Research Scientist, AllenNLP
 17. [03.2012] **Mahmoud Ghoneim, PhD**
Thesis: Explicit Modeling of Semantics in Statistical Machine Translation
 Joint Supervision with Mohsen Rashwan
 Computer Engineering Department, Faculty of Engineering, Cairo University, Egypt
 Current Position: Research Scientist, British Petroleum, Dallas Texas.
 18. [07.2009] **Yassine Benajiba, PhD**
Thesis: Arabic Named Entity Recognition
 Joint Supervision with Paolo Rosso
 Polytechnical University of Valencia, Spain
 Current Position: Research Scientist Manager, Amazon AWS AI, New York NY, USA

Current PhD Students [Date Started]

[Total: 3] (Date Started)

1. [Fall 2016] **Ali Seyfi**
Thesis Topic: Event Detection and Classification
 Department of Computer Science, The George Washington University, DC, USA

2. [Spring.2017] **Amal Qahtani**
Thesis Topic: Predictive models for mental illness
 Department of Computer Science, The George Washington University, DC, USA
3. [Fall.2017] **Rana Salama**
Thesis Topic: Text encoding using Wavelets
 Department of Computer Science, The George Washington University, DC, USA
 Joint Supervision with Dr. Abdou Youssef

PHD THESIS COMMITTEE MEMBERSHIP

[Total: 25]

1. [10.2022] **Christina Basta**
Thesis: Gender Bias in Machine Translation.
 Computer Science Department, Universitat Politècnica de Catalunya - Barcelona Tech, Barcelona, Spain.
2. [05.2022] **Chatrine Qwaider**
Thesis: Resources and Applications for Dialectal Arabic: The case for Levantine Arabic.
 Computer Science Department, Gottenberg University, Sweden.
3. [03.2022] **Jennifer Brooks**
Thesis: Automated Generation and Detection of Novel On-Topic Metaphors Using Machine Learning.
 Department of Computer Science, The George Washington University, DC, USA
4. [03.2022] **Zhantong Liang**
Thesis: Deep Sentence Denoising beyond Grammatical Error Correction.
 Department of Computer Science, The George Washington University, DC, USA
5. [12.2021] **Rehab Alahmadi**
Thesis: Learning Visual Features from Language to improve Image Captioning Performance
 Department of Computer Science, The George Washington University, DC, USA
6. [11.2021] **Bashar Alhafni**
Thesis Proposal: Natural language generation and gender bias in NLP, user-aware models in the context of morphologically rich languages, with a particular focus on Arabic.
 Department of Computer Science, NYU, Abu Dhabi, UAE
7. [07.2021] **Rehab Alahmadi**
Thesis Proposal: Learning Visual Features from Language to improve Image Captioning Performance
 Department of Computer Science, The George Washington University, DC, USA
8. [05.2021] **Elisbeth Turcan**
Thesis Proposal: Understanding Negative Emotional Reactions and Their Triggering Events: Interpretable Multitask Models
 Department of Computer Science, Columbia University, NYC USA
9. [04.2021] **Zhantong Liang**
Thesis Proposal: Deep Sentence Denoising beyond Grammatical Error Correction
 Department of Computer Science, The George Washington University, DC, USA
10. [03.2021] **Jennifer Brooks**
Thesis Proposal: Novel Metaphor Generation
 Department of Computer Science, The George Washington University, DC, USA
11. [05.2020] **Gustavo Aguilar**
Thesis: Neural Sequence Labeling on Social Media Text
 Department of Computer Science, University of Houston, Texas, USA
12. [04.2020] **Christopher Hidey**

- Thesis: Content Selection for Effective Counter-Argument Generation*
Department of Computer Science, Columbia University, NYC, USA
13. [10.2019] **Kazem Qazenfari**
Thesis: Combining Global and Local Knowledge for Text Analytics
Department of Computer Science, The George Washington University, DC, USA
14. [6.2019] **Noura Farra**
Thesis: Cross lingual Sentiment Analysis
Department of Computer Science, Columbia University, NYC, USA
15. [Summer.2017] **Bryan Wilkinson**
Thesis: Identifying and Ordering Scalar Adjectives Using Lexical Substitution
Department of Computer Science, University of Maryland Baltimore County, MD, USA
16. [Fall.2016] **Gilbert Badaro**
Thesis: Resources and Analytics for Opinion Mining and Recommender Systems, with Application to Arabic
Faculty of Engineering and Architecture, American University in Beirut (AUB), Lebanon
17. [Spring.2016] **Alawya Alawami**
Thesis: Aspect Extraction for Sentiment Analysis in Arabic Dialects
School of Information Sciences, University of Pittsburg, PA, USA
18. [Fall.2015] **Aous Mansouri**
Thesis: Stative and Stativizing Constructions in Arabic News Reports: A corpus based study
Department of Linguistics, University of Colorado, Boulder CO, USA
19. [Fall.2014] **Walaa Medhat**
Thesis: Text Mining on Social Networking using NLP Techniques
Faculty of Engineering, Ain Shams University, Cairo, Egypt
20. [Fall.2013] **John Sipple**
Thesis Proposal: Developing a language-learning model to incrementally infer syntactic structures from a continuous stream of English-language discourse.
Department of Computer Science, George Washington University, Washington DC, USA
21. [Spring.2012] **Aous Mansouri**
Thesis Proposal: Imperfective Constructions in Arabic.
Department of Linguistics, University of Colorado, Boulder CO, USA
22. [Spring.2010] **Daniele Pighin**
Thesis: Reverse Feature Engineering for SVMs
Computer Science Department, University of Trento, Italy
23. [Fall.2009] **Kapil Thadani**
Thesis Proposal: Semantic Fusion and Generation.
Department of Computer Science, Columbia University, New York NY, USA
24. [Spring.2009] **Bonaventura Coppola**
Thesis: Automatic Processing of Language Frame Semantics.
ICT Division, University of Trento, Italy
25. [Spring.2008] **Fadoua Ataa Allah**
Thesis: Information Retrieval: English and Arabic Documents.
Faculty of Science, University of Mohamed V, Rabat, Morocco

PUBLICATIONS

BOOKS

1. **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 2023.*

THESIS

1. **Diab, Mona.** (2003) Word Sense Disambiguation within a multilingual framework. *PhD Thesis, University of Maryland, College Park MD, USA.*

PATENTS

1. Granted Feb 14, 2023: Context aware Natural Language Understanding Models #**11580968**

EDITED CONFERENCE/WORKSHOP PROCEEDINGS

[Total: 13]

1. Tamar Solorio, Shuguang Chen, Alan W Black, **Mona Diab**, Sunayana Sitaram, Victor Soto, Emre Yilmaz; *Editors* (2021) Proceedings of the 5th Workshop on Computational Approaches to Linguistic Code-Switching, NAACL 2021, virtual
2. Parminder Bhatia, Steven Lin, Rashmi Gangadharaiyah, Byron C Wallace, Izhak Shafran, Chaitanya Shivade, Nan Du, **Mona Diab**; *Editors* (2020) Proceedings of the First Workshop on Natural Language Processing for Medical Conversations. ACL 2020, virtual
3. Tamar Solorio, Monojit Choudhury, Kalika Bali, Sunayana Sitaram, Amitava Das, and **Mona Diab**; *Editors* (2020) Proceedings of the 4th Workshop on Computational Approaches to Linguistic Code-Switching, LREC 2020, virtual
4. Gustavo Aguilar, Fahad Alghamdi, Victor Soto, Tamar Solorio, **Mona Diab**, Julia Hirshberg; *Editors.* (2018) ACL 2018 Proceedings of 3rd Workshop on Computational Approaches to Linguistic Code Switching (CALCS), ACL 2018, Melbourne, Australia.
5. Habash, Nizar, **Mona Diab**, Kareem Darwish, Wassim ElHajj, Hend Alkhalifa, Houda Bouamor; *Editors.* (2017) 2nd Workshop on Arabic Natural Language Processing (WANLP). In Proceedings of European chapter of the Association of Computational Linguistics (EACL).
6. **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).
7. **Diab, Mona**, Mahmoud Ghoneim, Houda Bouamor, Ahmed ElKholy, Yuval Marton, *Editors.* (2016) Workshop on Machine Translation for Semitic Languages (SEMAT). In Proceedings of Automatic Machine Translation in the Americas (AMTA).
8. **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).
9. **Diab, Mona**, Timothy Baldwin, Marco Baroni; *Editors.* (2013) 2nd International Joint Conference on Semantics (*SEM). Proceedings of *SEM 2013.
10. Agirre, Eneko, Johan Bos, **Mona Diab**; *Editors.* (2012) 1st International Joint Conference on Semantics (*SEM). Proceedings of *SEM 2012.
11. 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.
12. Choukri, Khalid, **Mona Diab**, Bente Maegaard, Paolo Rosso, Abdelhadi Soudi, 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.
13. Darwish, Kareem, **Mona Diab**, Nizar Habash, *Editors.* (2005) Proceedings of Computational Approaches to Semitic Languages. Association for Computational Linguistics.

JOURNAL ARTICLES

[Total: 10]

1. Bashir, Muhammad Huzaifa, Aqil M. Azmi, Haq Nawaz, Wajdi Zaghouni, **Mona Diab**, Ala AlFuqaha, Junaid Qadir. (2022) Arabic Natural Language Processing for Qur'anic Research: A Systematic Review. *Journal for Artificial Intelligence Review*.
2. Al Aqeel, Sinaa, Norah Abanmy, Abeer Aldayel, Hend Al-Khalifa, Maha Al-Yahya, and **Mona Diab**. (2018) Readability of written medicine information materials in Arabic language: expert and consumer evaluation. *BMC health services research* 18, no. 1 (2018): 139.
3. Aldarmaki, Hanan, Mahesh Mohan, **Mona Diab**. (2018) Unsupervised Word Mapping Using Structural Similarities in Monolingual Embeddings. *Transactions of the Association for Computational Linguistics (TACL)* 6, 185-196
4. Zaghouni, Wajdi, Abdelati Hawwari, **Mona Diab**. (2016) AMPN: A Lexical Semantic Resource for Arabic Morphological Patterns. *International Journal of Speech Technologies*, Springer Publishers.
5. 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)
6. Abdul-Mageed, Muhammad, and **Diab, Mona**. (2011) Linguistically-motivated subjectivity and sentiment annotation and tagging of Modern Standard Arabic. *International Journal on Social Media MMM: Monitoring, Measurement, and Mining*.
7. Katz, Graham, and **Mona Diab**. (2011) Introduction to the special issue on Arabic Computational Linguistics. *ACM Transactions on Asian Language Information Processing (TALIP)*. Volume 10 Issue 1, Mar. Article (1).
8. 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), pp.926-934.
9. 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), pp.464-473.
10. Resnik, Philip, Mari B Olsen, **Mona Diab**. (1999) The Bible as a Parallel Corpus: Annotating the 'Book of 2000 Tongues' *Computers and the Humanities*, 33. (201-2).

INVITED ARTICLES

[Total: 8]

1. **Diab, Mona**. (2015) Tharawat: A Vision for a Comprehensive Resource for Arabic Computational Processing. *International Conference on Intelligent Text Processing and Computational Linguistics*, Springer International Publishing, pp. 85-97.
2. 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.
3. **Diab, Mona** and Yuval Marton. (2014) Semantic Processing of Semitic Languages. *Book Chapter in Natural Language Processing for Semitic Languages*, Imed Zitouni (ed.), Springer Berlin Heidelberg, pp. 129-159.
4. Bar, Kfir, **Mona Diab**, Abdelati Hawwari. (2014) Arabic Multiword Expressions: Resource and Tool Creation. *Language Culture and Computation. Studies in honor of Yaacov Choueka. Naachum Derschowitz and Ephraim Nissim (Editors)*, *Lecture Notes in Computer Science, Vol.8001*, Springer-Verlag, Berlin, Germany.
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 Hacioglu, Daniel Jurafsky. (2007) Automatic Processing of Modern Standard Arabic Text. In *Arabic Computational Morphology: Knowledge-based and Empirical Methods*. Editors Antal van den Bosch and Abdelhadi Soudi. Kluwer/Springer Publications.

ARXIV ARTICLES (not published elsewhere yet)

[Total: 10]

1. Susan Zhang, Stephen Roller, Naman Goyal, Mikel Artetxe, Moya Chen, Shuohui Chen, Christopher Dewan, **Mona Diab**, Xian Li, Xi Victoria Lin, Todor Mihaylov, Myle Ott, Sam Shleifer, Kurt Shuster, Daniel Simig, Punit Singh Koura, Anjali Sridhar, Tianlu Wang, Luke Zettlemoyer. (2022) Opt: Open pre-trained transformer language models. *arXiv:2205.01068*
2. AlKhamissi, Badr, Millicent Li, Asli Celikyilmaz, **Mona Diab**^{*}, Maryam Ghazvininejad^{*}. (2022) A review on language models as knowledge bases. *arXiv:2204.06031* (* equal senior mentorship)
3. S Chen, G Aguilar, A Srinivasan, **M Diab**, T Solorio. (2022) CALCS 2021 Shared Task: Machine Translation for Code-Switched Data. *arXiv:2202.09625*
4. Peter Hase, **Mona Diab**, Asli Celikyilmaz, Xian Li, Zornitsa Kozareva, Veselin Stoyanov, Mohit Bansal, Srinivasan Iyer. (2021) Do Language Models Have Beliefs? Methods for Detecting, Updating, and Visualizing Model Beliefs. *arXiv:2111.13654*
5. Tamar Solorio, Mahsa Shafaei, Christos Smailis, **Mona Diab**, Theodore Giannakopoulos, Heng Ji, Yang Liu, Rada Mihalcea, Smaranda Muresan, Ioannis Kakadiaris. (2021) White Paper: Challenges and Considerations for the Creation of a Large Labelled Repository of Online Videos with Questionable Content. *arXiv:2101.10894*
6. P Hosseini, DA Broniatowski, **M Diab**. (2021) Predicting Directionality in Causal Relations in Text. *arXiv:2103.13606*
7. Alexander R. Fabbri, Xiaojian Wu, Srini Iyer, **Mona Diab**. (2021) Multi-Perspective Abstractive Answer Summarization. *arXiv:2104.08536*
8. Aditi Chaudhary, Siddharth Dalmia, Junjie Hu, Xinjian Li, Austin Matthews, Aldrian Obaja Muis, Naoki Otani, Shruti Rijhwani, Zaid Sheikh, Nidhi Vyas, Xinyi Wang, Jiateng Xie, Ruochen Xu, Chunting Zhou, Peter J Jansen, Yiming Yang, Lori Levin, Florian Metze, Teruko Mitamura, David R Mortensen, Graham Neubig, Eduard Hovy, Alan W Black, Jaime Carbonell, Graham V Horwood, Shabnam Tafreshi, **Mona Diab**, Efsun S Kayi, Noura Farra, Kathleen McKeown. (2019) The ARIEL-CMU Systems for LoReHLT18. *arXiv:1902.08899*
9. Prabhakaran, Vlnodkumar, Michael Bloodgood, **Mona Diab**, Bonnie Dorr, Lori Levin, Christine D. Piatko, Owen Rambow, Benjamin Van Durme. (2015) Statistical Modality Tagging from Rule Based Annotations and Crowdsourcing. *arXiv:1503.01190*
10. **Diab, Mona**, Nizar Habash, Owen Rambow, Ryan Roth. (2013) LDC Arabic Treebanks and Associated Corpora: Data Divisions Manual. *arXiv:1309.5652*

PEER REVIEWED CONFERENCE PAPERS

[Total: 113]

[2023]

1. Ping Yu, Tianlu Wang, Olga Golovneva, Badr Alkhamissy, Gargi Ghosh, **Mona Diab**, Asli Celikyilmaz. (2023) ALERT: Adapting Language Models to Reasoning Tasks. *ACL 2023*
2. Peter Hase, **Mona Diab**, Asli Celikyilmaz, Xian Li, Zornitsa Kozareva, Veselin Stoyanov, Mohit Bansal and Srinivasan Iyer. (2023) Methods for Measuring, Updating, and Visualizing Factual Beliefs in Language Models. *EACL 2023*

[2022]

3. Bergman, Stevie A., **Mona Diab**. (2022) Towards Responsible Natural Language Annotation for the Varieties of Arabic. *ACL 2022 Findings*
4. Alex R Fabbri, Haoran Li, Xioajian Wu, Srini Iyer, **Mona Diab**. (2022) AnswerSumm: A Manually-Curated Dataset and Pipeline for Answer Summarization. *NAACL 2022*
5. Mikel Artetxe, Shruti Bhosale, Naman Goyal, Todor Mihaylov, Myle Ott, Sam Shleifer, Xi Victoria Lin, Jingfei Du, Srinivasan Iyer, Ramakanth Pasunuru, Giri Anantharaman, Xian Li, Shuohui Chen, Halil Akin, Mandeep Baines, Louis Martin, Xing Zhou, Punit Singh Koura, Brian O'Horo, Jeff Wang, Luke Zettlemoyer, **Mona Diab**, Zornitsa Kozareva, Ves Stoyanov. (2022). Efficient Large Scale Language Modeling with Mixtures of Experts. *EMNLP 2022*
6. Xi Victoria Lin, Todor Mihaylov, Mikel Artetxe, Tianlu Wang, Shuohui Chen, Daniel Simig, Myle Ott, Naman Goyal, Shruti Bhosale, Jingfei Du, Ramakanth Pasunuru, Sam Shleifer, Punit Singh Koura, Vishrav Chaudhary, Brian O'Horo, Jeff Wang, Luke Zettlemoyer, Zornitsa Kozareva, **Mona Diab**, Veselin Stoyanov, Xian Li. (2022) Few-shot Learning with Multilingual Language Models. *EMNLP 2022*
7. Jennifer Tracey, Owen Rambow, Claire Cardie, Adam Dalton, Hoa Trang Dang, **Mona Diab**, Bonnie Dorr, Louise Guthrie, Magdalena Markowska, Smaranda Muresan, Vinodkumar Prabhakaran, Samira Shaikh, Tomek Strzalkowski. (2022) BeST: The Belief and Sentiment Corpus. *LREC 2022*
8. Daniel Simig, Tianlu Wang, Verna Dankers, Peter Henderson, Khuyagbaatar Batsuren, Dieuwke Hupkes, **Mona Diab**. (2022) Text Characterization Toolkit [TCT]. *ACL 2022 Demo Track*
9. Badr Alkhamissi, Faisal Ladhak, Srini Iyer, Ves Stoyanov, Zornitsa Kozareva, Xian Li, Pascale Fung, Lambert Mathias, Asli Celikyilmaz, **Mona Diab**. (2022) ToKen: Task Decomposition and Knowledge Infusion for Few-Shot Hate Speech Detection. *EMNLP 2022 Oral*
10. Daniel Licht, Cynthia Gao, Janice Lam, Francisco Guzman, **Mona Diab***, Philipp Koehn*. (2022) Consistent Human Evaluation of Machine Translation across Language Pairs. *AMTA 2022 Oral*

[2021]

11. Parsa Farinneya, Mohammad Mahdi Abdollah Pour, Sardar Hamidian, **Mona Diab**. (2021) Active Learning for Rumor Identification on Social Media. *EMNLP Findings*.
12. N Almarwani, **M Diab**. (2021) Discrete Cosine Transform as Universal Sentence Encoder. *ACL Oral*

13. A Renduchintala, D Diaz, K Heafield, X Li, **M Diab**. (2021) [Gender Bias Amplification During Speed-Quality Optimization in Neural Machine Translation](#). *ACL Oral*
14. Wei-Jen Ko, Ahmed El-Kishky, Adithya Renduchintala, Vishrav Chaudhary, Naman Goyal, Francisco Guzmán, Pascale Fung, Philipp Koehn, **Mona Diab**. (2021) [Adapting High-resource NMT Models to Translate Low-resource Related Languages without Parallel Data](#). *ACL Oral*
15. Chunting Zhou, Jiatao Gu, Mona T. Diab, P. Guzmán, Luke Zettlemoyer, Marjan Ghazvininejad. (2021) [Detecting hallucinated content in conditional neural sequence generation](#). *ACL Findings*

[2020]

16. M Aminian, MS Rasooli, **M Diab**. (2020) [Multitask Learning for Cross-Lingual Transfer of Broad-coverage Semantic Dependencies](#). *EMNLP 2020*
17. Sawsan Alqahtani, Ajay Mishra, Mona Diab. (2020) [A multitask learning approach for diacritic restoration](#). *ACL 2020*
18. Esin Durmus, He He, **Mona Diab**. (2020) [FEQA: A question answering evaluation framework for faithfulness assessment in abstractive summarization](#). *ACL 2020*
19. Christopher Hidey, Tuhin Chakrabarty, Tariq Alhindi, Siddharth Varia, Kriste Krstovski, **Mona Diab**, Smaranda Muresan. (2020) [DeSePtion: Dual sequence prediction and adversarial examples for improved fact-checking](#). *ACL 2020*
20. Yi-An Lai, Xuan Zhu, Yi Zhang, **Mona Diab**. (2020) [Diversity, density, and homogeneity: Quantitative characteristic metrics for text collections](#). *LREC 2020*
21. Efsun Sarioglu Kayi, Linyong Nan, Bohan Qu, **Mona Diab**, Kathleen McKeown. (2020) [Detecting Urgency Status of Crisis Tweets: A Transfer Learning Approach for Low Resource Languages](#). *COLING 2020*

[2019]

22. Sawsan Alqahtani and **Mona Diab**. (2019) [Investigating Input and Output Units in Diacritic Restoration](#). *Proceedings of IEEE ICMLA, Dec. 16-19, Boca Raton, Florida, USA*.
23. Amal Alqahtani, Efsun Kayi, **Mona Diab**. (2019) [Understanding Cohesion in Writings of Schizophrenia Patients](#). *Proceedings of IEEE ICMLA, Dec. 16-19, Boca Raton, Florida, USA*.
24. Almarwani, Nada; Hanan Aldarmaki, **Mona Diab**. (2019) [Efficient Sentence Embedding using Discrete Cosine Transform](#). *Proceedings of EMNLP-IJCNLP 2019, Hong Kong, Nov. 2019*.
25. Alqahtani, Sawsan; Ajay Mishra, **Mona Diab**. (2019) [Convolutional Neural Networks for Diacritic Restoration](#). *Proceedings of EMNLP-IJCNLP 2019, Hong Kong, Nov. 2019*.
26. Gupta, Arshit; Peng Zhang, Garima Lalwani, **Mona Diab**. (2019) [CASA-NLU: Context-Aware Self-Attentive Natural Language Understanding for Task-Oriented Chatbots](#). *Proceedings of EMNLP-IJCNLP 2019, Hong Kong, Nov. 2019*.
27. Peskov, Denis; Nancy Clarke, Jason Krone, Brigi Fodor, Yi Zhang, Adel Youssef, **Mona Diab**. (2019). [Multidomain Goal Oriented Dialogues \(MultiDoGo\): Strategies toward the curation and Annotation of large scale dialogue datasets](#). *Proceedings of EMNLP-IJCNLP 2019, Hong Kong, Nov. 2019*.
28. Hosseini, Pedram, **Diab, Mona**, and David Broniatowski. (2019) [Does Causal Coherence predict online spread of social media?](#) *In Proceedings of International Conference on Social Computing, Behavioral-Cultural Modeling and Production and Behavior Representation in Modeling and Simulation (SBP-BRIMS)*, Washington, DC USA, July 2019.
29. Aldarmaki, Hanan, and **Mona Diab**. (2019) [Context Aware Cross Lingual Mapping](#). *Proceedings of NAACL 2019, Minneapolis, MN, USA, June 2019*.
30. Aldarmaki, Hanan, and **Mona Diab**. (2019) [Scalable Cross-Lingual Transfer of Neural Sentence Embeddings](#). *Proceedings of *SEM 2019, Minneapolis, MN, USA, June 2019*.

31. Aminian, Maryam, Mohamed Sadegh Rasooli, and **Mona Diab**. (2019) Cross-Lingual Transfer of Semantic Roles: From Raw Text to Semantic Roles. *Proceedings of 13th International Conference on Computational Semantics (IWCS 2019)*, May, Gottenburg, Sweden.

[2018]

32. W. Zaghouani, S. Alqahtani, **M. Diab**. (2018) Building a Rich Lexical Resource for Standard Arabic. *Qatar Foundation Annual Research Conference Proceedings*, Qatar, October 2018.
33. Hosseini, Pedram, **Diab, Mona**, and David Broniatowski. (2018) False news in social media? *In Proceedings of International Conference on Social Computing, Behavioral-Cultural Modeling and Production and Behavior Representation in Modeling and Simulation (SBP-BRiMS)*, Washington, DC USA, July 2018.
34. Hosseini, Pedram, **Diab, Mona**, and David Broniatowski. (2018) A spread of Coherent Stories on social media? *In Proceedings of 2nd North American Social Networks Conference NASN 2018*, Washington, DC USA, November 2018.
35. Alghamdi, Fahad, **Mona Diab**. (2018) WASA: A Web Application for Sequence Annotation. *Proceedings of LREC 2018. Miyazaki, Japan, May 2018*.
36. Tafreshi, Shabnam, **Mona Diab**. (2018) Sentence and Clause Level Emotion Annotation, Detection, and Classification in a Multi-Genre Corpus. *Proceedings of LREC 2018. Miyazaki, Japan, May*.
37. Hanan Aldarmaki and **Mona Diab**. (2018) Evaluation of Unsupervised Compositional Representations. *COLING 2018*, New Mexico, USA, August 2018.
38. Shabnam Tafreshi and **Mona Diab**. (2018) Emotion Detection and Classification in a Multigenre Corpus with Joint Multi-Task Deep Learning. *COLING 2018*, New Mexico, USA, August 2018.

[2017]

39. Aminian, Maryam, Mohamed Sadegh Rasooli, **Mona Diab**. (2017) Transferring Semantic Roles Using Translation and Syntactic Information. *Proceedings of IJCNLP 2017*, Taipei Taiwan, December.
40. Efsun Kayi, **Mona Diab**, Luca Pauselli, Michael Compton, Glen Coppersmith. (2017) Predictive Linguistic Features of Schizophrenia. *Proceedings of *SEM 2017, Vancouver Canada*, August.

[2016]

41. AlBadrashiny, Mohamed and **Mona Diab**. (2016) LILI: A Simple Efficient Language Independent Framework for Linguistic Code Switch Point Detection. *Proceedings of COLING 2016, Japan*, Dec.
42. AlQahtani, Sawsan, Mahmoud Ghoneim, **Mona Diab**. (2016) Impact of Explicit encoding of vowelization in Arabic Machine translation. *Proceedings of Automatic Machine Translation Association (AMTA) 2016, Texas Austin, USA, Nov*.
43. 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*.
44. **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*.
45. Zaghouani, Wajdi, Houda Bouamor, Abdelati Hawwari, **Mona Diab**, Ossama Obeid, Mahmoud Ghoneim, Sawsan Alqahtani, Kemal Oflazer. (2016) Guidelines and Framework for a Large Scale Arabic Diacritized Corpora. *In Proceedings of LREC 2016, Slovenia, May*.
46. Al-Badrashiny, Mohamed, Arfath Pasha, **Mona Diab**, Nizar Habash, Owen Rambow, Wael

Salloum, and Ramy Eskander, (2016) [SPLIT: Smart Preprocessing \(Quasi\) Language Independent Tool](#), *In Proceedings of LREC 2016, Slovenia, May.*

[2015]

47. Prabhakaran, Vinodkumar, 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.
48. Elfardy, Heba, **Mona Diab**, Chris Callison-Burch. (2015) [Ideological Perspective Detection Using Semantic Features](#). *In Proceedings of 4th *SEM Conference*, Denver CO, USA.
49. 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.
50. Hamidian, Sardar, **Mona Diab**. (2015) [Rumor Detection and Classification for Twitter Data](#). *In Proceedings of The Fifth International Conference on Social Media Technologies, Communication, and Informatics (SOTICS)*, Barcelona, Spain.
51. Zirikly, Aya, Yassine Benajiba, **Mona Diab**. (2015) [GW-EDL English Diagnostic Short System Description](#). KPB Task Evaluation, *Proceedings of the TAC Conference*, NIST, Maryland MD, USA.

[2014]

52. 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.
53. **Diab, Mona**, Mohamed AlBadrashiny, Maryam Aminian, Mohamed Attia, Hebatullah Elfardy, 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.
54. Guo, Weiwei, Wei Liu, **Mona Diab**. (2014) [Fast Tweet Retrieval with Compact Binary Codes](#). *In Proceedings of Conference of Computational Linguistics (COLING)*, Dublin, Ireland.
55. Pasha, Arfath, Mohamed AlBadrashiny, **Mona Diab**, Ahmed ElKholly, 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.
56. Salloum, Wael, Hebatullah Elfardy, 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.

[2013]

57. 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.
58. 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.
59. 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.

60. ElFardy, Hebatullah, **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.
61. ElFardy, 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.
62. Ghoneim, Mahmoud and **Mona Diab**. (2013) Multiword Expressions in the context of Statistical Machine Translation. *In Proceedings of IJCNLP*, Nagoya, Japan.
63. 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.
64. Guo, Weiwei, Hao Li, Heng Ji, **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.
65. 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.
66. Tomeh, Nadi, Nizar Habash, Ryan Roth, Noura Farra, Pradeep Dasigi and **Mona Diab**. (2013) Reranking with Linguistic and Semantic Features for Arabic Optical Character Recognition, *In Proceedings of Association for Computational Linguistics (ACL)*, Sofia, Bulgaria.
67. 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.

[2012]

68. 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.
69. 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.
70. 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.
71. Dasigi, Pradeep, Weiwei Guo, and **Diab, Mona (2012)** Genre Independent Subgroup Detection in Online Discussion Threads: A Pilot Study of Implicit Attitude using Latent Textual Semantics. *In Proceedings of Association for Computational Linguistics (ACL)*, Jeju Island, South Korea.
72. 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.
73. 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.
74. Guo, Weiwei and **Mona Diab**. (2012) Modeling Sentences in the Latent Space. *In Proceedings of Association for Computational Linguistics (ACL)*, Jeju Island, South Korea.
75. 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.

76. 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.
77. 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.
78. 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.
79. 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. 8662476620, 59322389

[2011]

80. 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.
81. **Diab, Mona** and Pradeep Dasigi. (2011) CODACT: Towards Identifying Orthographic Variants in Dialectal Arabic. In *Proceedings of IJCNLP*, Chiang Mai, Thailand.
82. 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.
83. 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.

[2010]

84. 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.
85. 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.
86. 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.
87. Prabhakaran, Vinodkumar, Owen Rambow, **Mona Diab**. (2010) Automatic Committed Belief Tagging. In *Proceedings of COLING*, Beijing, China.

[2009]

88. **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.
89. **Diab, Mona** and Madhav Krishna. (2009) Unsupervised classification of Verb Noun Multiword Expression Tokens. *CICLING*, Mexico City, Mexico.
90. 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. (2009) What, When, Where, Why? Comparing Multiple Approaches to the Cross-Lingual 5W Task. In *Proceedings of ACL-IJCNLP*, Singapore.

[2008]

91. 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.
92. 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.
93. **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.
94. 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.
95. 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.

[2007]

96. **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.
97. **Diab, Mona** and Alessandro Moschitti. (2007) Semantic Parsing for Modern Standard Arabic. *In Proceedings of Recent Advances in Natural Language Processing (RANLP)*, Borovets, Bulgaria.
98. **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.
99. 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.

[2006]

100. 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.
101. **Diab, Mona**, and Neal Snider. (2006) Using the Arabic Treebank for Lexical Semantics. (Abstract and Presentation). *In Proceedings of Annual Meeting of the Arabic Linguistics Society (ALS)*, Kalamazoo MI, USA.
102. 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.
103. Snider, Neal and **Mona Diab**. (2006) Unsupervised Induction of Arabic Verb Classes using Syntactic Frames and LSA. *In Proceedings of the Joint Conference of the International Committee on Computational Linguistics and the Association for Computational Linguistics. (ACL-COLING)*, Sydney, Australia.
104. 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.

[2004]

105. **Diab, Mona.** (2004) Relieving the Data Acquisition Bottleneck for Word Sense Disambiguation. *In Proceedings of Association for Computational Linguistics (ACL).*
106. **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).*
107. **Diab, Mona.** (2004) Feasibility of Bootstrapping a WordNet Taxonomy for Arabic Leveraging Parallel Corpora and an English WordNet. *Proceedings of First Arabic Language Technology Conference. (NEMLAR), Cairo, Egypt.*

[2002]

108. **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.*

[2000]

109. **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 (RIA000), Paris, France.*
110. Resnik, Philip and **Mona Diab.** (2000) Measuring Verb Similarity. *In Proceedings of the conference of International Cognitive Science Society (COGSCI00), Philadelphia PA, USA.*

[1998]

111. **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.*
112. 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.*

[1997]

113. 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

[Total: 76]

[2023]

1. Alkhamissi, Badr, S. Verma, P. Yu, Z. Jin, A. Celikyilmaz, M. Diab. (2023) OPT-R: Exploring the Role of Explanations in Finetuning and Prompting for Reasoning Skills of Large Language Models. *Proceedings of Workshop on Natural Language Reasoning and Structured Explanations, ACL 2023, Toronto, Canada*

2. Y Bang, T Yu, A Madotto, Z Lin, **M Diab**, P Fung. (2023) Enabling Classifiers to Make Judgements Explicitly Aligned with Human Values. Proceedings of 3rd Workshop on Trustworthy NLP (*TrustNLP*), *ACL 2023, Toronto, Canada*

[2022]

3. Hosseini, Pedram, Christopher R. Wolfe, **Mona Diab**, David A. Broniatowski. (2022) GisPy: A Tool for Measuring Gist Inference Score in Text. *Proceedings of the 4th workshop on Narrative Understanding, NAACL 2022*
4. Pedram Hosseini, David A. Broniatowski, **Mona Diab**. (2022) Knowledge-Augmented Language Models for Cause-Effect Relation Classification. *Proceedings of the First Workshop on Commonsense Representation and Reasoning (CSRR 2022), ACL 2022*
5. Alqahtani, Amal, Efsun S Kayi, Sardar Hamidian, Michael Compton, **Mona Diab**. (2022) A Quantitative and Qualitative Analysis of Schizophrenia Language. *Proceedings of LOUHI 2022 Workshop, EMNLP 2022*
6. **Badr AlKhamissi, Mona Diab**. (2022) Meta AI at Arabic Hate Speech 2022: MultiTask Learning with Self-Correction for Hate Speech Classification. *Proceedings of OSACT Workshop, LREC 2022*. **[BEST PAPER AWARD]**
7. Mohammed EINokrashy, Badr AlKhamissi, **Mona Diab**. (2022) Depth-Wise Attention (DWAtt): A Layer Fusion Method for Data-Efficient Classification. *Proceedings of ENLSP Workshop, NeurIPS 2022*

[2020]

8. Jason Krone, Yi Zhang, **Mona Diab**. (2020) Learning to Classify Intents and Slot Labels Given a Handful of Examples. *Proceedings of Workshop on Conversational AI, ACL 2020*

[2019]

9. Christopher Hidey, Tuhin Chakrabarty, Tariq Alhindi, Siddharth Varia, Kriste Krstovski, **Mona Diab**, Smaranda Muresan. (2019) Cure My FEVER: Building, Breaking, and Fixing Models for Fact Checking. *Proceedings of EMNLP 2nd FEVER Workshop, Hong Kong, November 2019*.
10. Or Levi, Pedram Hosseini, **Mona Diab**, David Broniatowski. (2019) Identifying nuances in Fake News vs. Satire: Using Semantic and Linguistic Cues. *Proceedings of EMNLP 2nd Workshop on NLP for Internet Freedom (NLP4IF): Censorship, Disinformation, and Propaganda, Hong Kong, November 2019*.
11. **Sawsan Alqahtani, Hanan Aldarmaki, Mona Diab**. (2019) Homograph Disambiguation via Selective Diacritic Restoration. *Proceedings of the WANLP 2019, ACL2019, Florence, Italy, July 2019*.
12. Hamidian, Sardar and **Mona Diab**. (2019) GWU NLP at SemEval-2019 Task 7: Hybrid Pipeline for Rumour Veracity and Stance Classification on Social Media. *Proceedings of SemEval 2019, NAACL 2019, Minnesota, MN, USA, June 2019*.
13. Tafreshi, Shabnam, and **Mona Diab**. (2019) GWU NLP at SemEval-2019 Task 3: EmoContext: Effective Contextual Information in Models for Emotion Detection in Sentence-level in a Multigenre Corpus. *Proceedings of SemEval 2019, NAACL 2019, Minnesota, MN, USA, June 2019*.
14. **AlGhamdi, Fahad, and Mona Diab**. (2019) Leveraging Pretrained Word Embeddings for Part-of-Speech Tagging of Code Switching Data. *Proceedings of VarDial 2019, NAACL 2019, Minnesota, MN, USA, June 2019*.

[2018]

15. Sawsan Alqahtani, **Mona Diab** and Wajdi Zaghrouani. (2018) ARLEX: A Large Scale Comprehensive Lexical Inventory for Modern Standard Arabic. *Proceedings of the Workshop on Open-Source Arabic Corpora and Processing Tools (OSACT3)*, LREC 2018, Miyazaki, Japan.
16. Abeer AL-Dayel, Hend Al-Khalifa, Sinaa Alaqeel, Norah Abanmy, Maha Al-Yahya and **Mona Diab**. (2018) ARC-WMI: Towards Building Arabic Readability Corpus for Written Medicine Information. *Proceedings of the Workshop on Open-Source Arabic Corpora and Processing Tools (OSACT3)*, LREC 2018, Miyazaki, Japan.
17. Christopher Hidey and **Mona Diab**. (2018) Team SWEEPer: Joint Sentence Extraction and Fact Checking with Pointer Networks. Pp. 150-155. *Proceedings of the First Workshop on Fact Extraction and VERification (FEVER)*, EMNLP 2018, Brussels, Belgium, October 2018.
18. G Aguilar, F AlGhamdi, V Soto, **M Diab**, J Hirschberg, T Solorio. (2018) Named Entity Recognition on Code Switched Data: Overview of the CALCS 2018 Shared Task. *Proceedings of the 3rd Workshop on Computational Approaches to Linguistic Code Switching*, ACL 2018, Melbourne, Australia.

[2017]

19. Al-Badrashiny, Mohamed, Abdelati Hawwari, **Mona Diab**. (2017) A layered language model based hybrid approach to automatic full diacritization of Arabic. *Proceedings of EACL Workshop on Arabic language processing*, EACL 2017, Valencia Spain, Feb.
20. Nada Almarwani, **Mona Diab**. (2017) Arabic Textual Entailment with Word Embeddings. *Proceedings of EACL Workshop on Arabic language processing*, EACL 2017, Valencia Spain, Feb.
21. Nada Almarwani, **Mona Diab**. (2017) GW QA at SemEval-2017 Task 3: Question Answer Re-ranking on Arabic Fora. *Proceedings of ACL Workshop on Semantic Evaluation (SEM-EVAL 2017)*, ACL 2017, Vancouver, BC, Canada, August.
22. Daniel Cer, **Mona Diab**, Eneko Agirre, Inigo Lopez-Gazpio, Lucia Specia. (2017) SemEval-2017 Task 1: Semantic Textual Similarity Multilingual and Crosslingual Focused Evaluation. *In Proceedings of SEM-EVAL 2017, Association for Computational Linguistics*, ACL 2017, Vancouver, Canada, August

[2016]

23. Al-Badrashiny, Mohamed, and Abdelati Hawwari, Mahmoud Ghoneim, **Mona Diab**. (2016) SAMER: A Semi-Automatically Created Lexical Resource for Arabic Verbal Multiword Expressions Tokens Paradigm and their Morphosyntactic Features. *Proceedings of COLING 12th Workshop on Asian Language Resources (ALR12)*, COLING 2016, Osaka, Japan, Dec.
24. Aminian, Maryam, Mohamed Al-Badrashiny, **Mona Diab**. (2016) Automatic Verification and Augmentation of Multilingual Lexicons. *Proceedings of COLING Third Workshop on NLP for Similar Languages, Varieties and Dialects (VARDIAL3)*, COLING 2016, Osaka, Japan, Dec.
25. Attia, Mohammed, and Ayah Zirikly, **Mona Diab**. (2016) The Power of Language Music: Arabic Lemmatization through Patterns. *Proceedings of COLING Workshop on Cognitive Aspects of the Lexicon (CogALex-V)*, COLING 2016, Osaka, Japan, Dec.
26. Zaghrouani, Wajdi, Abdelati Hawwari, Sawsan Alqahtani, Houda Bouamor, Mahmoud Ghoneim, **Mona Diab**, Kemal Oflazer. (2016) Using Ambiguity Detection to Streamline Linguistic Annotation. *Proceedings of COLING Workshop on Computational Linguistics for Linguistic Complexity (CL4LC)*, COLING 2016, Osaka, Japan, Dec.
27. Zaghrouani, Wajdi, Nizar Habash, Houda Bouamor, Ossama Obeid, Sawsan Alqahtani, **Mona Diab**, Kemal Oflazer. (2016) Filtering Dialectal Arabic Text in Two Large Scale Annotation Projects. *Proceedings of COLING 2nd Workshop on 2016 The 2nd Workshop on Noisy User-generated Text (W-NUT)*, COLING 2016, Osaka, Japan, Dec.

28. Zirikly, Ayah, Bart Desmet, **Mona Diab**. (2016) The GW/LT3 VarDial 2016 Shared Task System for Dialects and Similar Languages Detection. *Proceedings of COLING Third Workshop on NLP for Similar Languages, Varieties and Dialects (VARDIAL3)*, COLING 2016, Osaka, Japan, Dec.
29. Molina, Giovanni, Fahad Algamdi, Mahmoud Ghoneim, Abdelati Hawwari, **Mona Diab**, Solorio, Tamar. (2016) An overview of the shared task on multilingual linguistic code switching. *Proceedings of EMNLP Workshop on Computational Approaches to Linguistic Code Switching (CALCS 2016)*, EMNLP, Austin TX, USA, Nov.
30. Alghamdi, Fahad, Giovanni Molina, **Mona Diab**, Tamar Solorio. (2016) Part of Speech Taggers for Code Switched Data. *Proceedings of EMNLP Workshop on Computational Approaches to Linguistic Code Switching (CALCS 2016)*, EMNLP, Austin TX, USA, Nov.
31. AlBadrashiny, Mohamed and **Mona Diab**. (2016) The George Washington University System for the Code-Switching Workshop Shared Task 2016. *Proceedings of EMNLP Workshop on Computational Approaches to Linguistic Code Switching (CALCS 2016)*, EMNLP, Austin TX, USA, Nov.
32. 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. **BEST POSTER AWARD FROM QNRF**
33. Abdul-Mageed, Muhammad, Hassan AlHuzli, 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.
34. Elfardy, Heba and **Mona Diab**. (2016) CU-GW 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.
35. Elfardy, Heba and **Mona Diab**. (2016) Addressing 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
36. AlDarmaki, Hanan, and **Mona Diab**. (2016) GW-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.
37. 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.
38. Hamidian, Sardar, 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.
39. 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

[2015]

40. 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.
41. 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.***

42. 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.
43. 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.
44. 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.
45. Attia, Mohammed, Mohamed Al-Badrashiny, **Mona Diab**. (2015) GW-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.
46. 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

[2014]

47. 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.
48. 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.
49. Attia, Mohamed, Mohamed AlBadrashiny, **Mona Diab**. (2014) GW-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.
50. EIFardy, 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.
51. 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.
52. 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.
53. 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.

[2012]

54. Abdul-Mageed, M., Kuebler, S., **Mona Diab**. (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.
55. 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.
56. 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.
57. **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.
58. Prabhakaran, Vinodkumar, Michael Bloodgood, **Mona Diab**, Bonnie J. Dorr, Lori Levin, Christine Piatko, Owen Rambow, 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.
59. Zaghouni, Wajdi, Abdelati Hawwari, **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.

[2011]

60. 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.
61. Dasigi, Pradeep and **Mona Diab**. (2011) Named Entity Transliteration Generation Leveraging Statistical Machine Translation Technology. In *Proceedings of NEWS 11 Workshop, IJCNLP, Chiang Mai, Thailand*.
62. **Diab, Mona** and Ankit Kamboj. (2011) Feasibility of Leveraging Crowd Sourcing for the Creation of a Large Scale Annotated Resource for Hindi English Code Switched Data: A Pilot Annotation. In *Proceedings of ALR9 Workshop, IJCNLP, Chiang Mai, Thailand*

[2010]

63. **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.
64. 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
65. Guo, Weiwei and **Mona Diab**. (2010) COLEUR and COLSLM: A WSD approach to Multilingual Lexical Substitution, Tasks 2 and 3 SemEval 2010. In *Proceedings of SEMEVAL, Association for Computational Linguistics (ACL)*, Uppsala, Sweden.
66. 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.

[2009]

67. **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.
68. **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.
69. **Diab, Mona** and Madhav Krishna. (2009) Handling Sparsity for Unsupervised MWE Classification. *GEMS Workshop, European Association for Computational Linguistics (EACL)*, Athens, Greece.
70. 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.

[2008]

71. 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.

[2007]

72. **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.
73. **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.
74. **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.

[2004]

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

[2000]

76. **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

[Total: 11]

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. Baker, Kathryn, Michael Bloodgood, **Mona Diab**, Bonnie J. Dorr, Ed Hovy, Lori Levin, Marjorie McShane, Teruko Mitamura, Sergei Nirenburg, Christine Piatko, Owen Rambow, Gramm Richardson (2010) SIMT SCALE 2009 - Modality Annotation Guidelines. Technical Report 004, Human Language Technology Center of Excellence, Johns Hopkins University, Baltimore, 2010.
 5. 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
 6. **Diab, Mona**. (2000) Exploiting Translations for Semantic Annotation. Candidacy Exam Thesis, Linguistics Department, University of Maryland, College Park MD, USA.
 7. 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
 8. 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
 9. **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
 10. **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
 11. **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.

PRESENTATIONS & POSTERS (Not Associated with Publications)

[Total: 40]

1. *Sentence level Emotion Detection and Classification using Portable Semantic features*. Shabnam Tafreshi and **Mona Diab**. (2016) *Women in Machine Learning Symposium, Spain*. December.
2. *Emotion, Intensity, and Urgency*. (2016) **DARPA LORELEI Site Visit**, Columbia University, New York NY, USA, Jul.
3. *Committed Belief Tagging*. (2016) **DARPA DEFT Site Visit**, Columbia University, New York NY, USA, Jul.
4. *Named Entity Linking & Event Detection*. (2016) **DARPA DEFT Site Visit**, Columbia University, New York NY, USA, Jul.

5. *Arabic Dialect Tools*. (2014) **DARPA BOLT Site Visit**, March, Columbia University, New York NY, USA, Mar.
6. *Perspective, Novelty and Belief in Streaming data and informal genres*. (2014) **DARPA DEFT Site visit**, George Washington University, Washington DC, USA, Feb.
7. *Arabic Dialect Processing under BOLT*. Invited Presentation. (2011) **DARPA BOLT Kick-off Meeting**, Tampa FL, USA, Oct.
8. *Arabic Language Modeling for OCR*. (With Nizar Habash) Invited Presentation. (2011) **DARPA MADCAT PI Meeting**. Boston MA, USA, Jul.
9. *Arabic Subjectivity and Sentiment Analysis: System & Resources*. Invited Poster. (2011) **DARPA BOLT Kick-off Meeting**, Tampa FL, USA, Oct.
10. *CADIM Arabic MSA NLP Tools*. (With Nizar Habash and Owen Rambow) Invited Poster, (2011) **DARPA BOLT Kick-off Meeting**, Tampa FL, USA, Oct.
11. *COLABA Tools and Resources*. (With Nizar Habash and Owen Rambow) Invited Presentation (2011) **DARPA BOLT Kick-off Meeting**, Tampa FL, USA, Oct.
12. *COLABA Tools and Resources*. (With Nizar Habash and Owen Rambow) Invited Presentation. (2011) **Brigham Young University**, Salt Lake City UT, USA, Aug 11.
13. *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.
14. *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.
15. *Power Shifts & Rifts in Social Communications*. (With Kathleen McKeown, Owen Rambow, Dragomir Radev) Invited Presentation. (2011) **IARPA SCIL PI Meeting**. Washington DC, USA, May.
16. *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.
17. *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.
18. *Arabic Language Modeling for OCR*. (With Nizar Habash) Invited Presentation. (2010) **DARPA MADCAT PI Meeting**. Boston MA, USA, Sep.
19. *COLABA Annotation, Tools, Resources*. (With Nizar Habash and Owen Rambow) Invited Presentation. (2010) **Government Day AMTA 10**. Denver CO, USA, Oct.
20. *Power Shifts & Rifts*. (With Kathleen McKeown, Owen Rambow, Dragomir Radev) Invited Presentation. (2010) **IARPA SCIL PI Meeting**. Washington DC, USA, May.
21. *Arabic Language Modeling for OCR*. (With Nizar Habash) Invited Presentation. (2009) **DARPA MADCAT PI Meeting**. Boston MA, USA, Sep.
22. *Power Shifts & Rifts*. (With Kathleen McKeown, Owen Rambow, Dragomir Radev) Invited Presentation. (2009) **IARPA SCIL PI Meeting**. Washington DC, USA, Oct.
23. *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.
24. *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.
25. *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.
26. *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.
27. *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.
 28. *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.
 29. *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.
 30. *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.
 31. *NLP Tools for Arabic*. (With Nizar Habash and Owen Rambow) Invited Presentation (2006) **DARPA GALE PI Meeting**. Boston MA, USA, May.
 32. *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.
 33. *An Unsupervised method for Word Sense Tagging using Parallel corpora*. (2001) Poster Session: **UMIACS Research Day, University of Maryland**. College Park MD, USA, Mar.
 34. *A Statistical Model of Word-Level Mapping for Comparable Corpora*. (2000) **UMIACS Computational Linguistics Colloquium, University of Maryland**. College Park MD, USA, Spr.
 35. *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.
 36. *MADLIBS: An LCS based Translingual Information Retrieval System*. (With Maria Katsova) (2000) **Demonstration: NAACL2000**. Seattle WA, USA, May.
 37. *Statistical Methods and Linguistics*. (With Elixabete Murguia) (2000) **LingFest, Linguistics Department University of Maryland**. College Park MD, USA, Nov.
 38. *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.
 39. *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.
 40. *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.

RELEASED SOFTWARE

RELEASED ROBUST TECHNOLOGY

1. [DIRA] Dialectal Arabic Information Retrieval Assistant. (2013)
Demo: <http://nlp.ldeo.columbia.edu/dira>
2. [AIDA] Automatic Identification of Dialectal Arabic. (2013-present)
Demo: <http://nlp.ldeo.columbia.edu/aida>

3. [MADAMIRA] A system for morphological analysis and disambiguation, base phrase chunking and named entity recognition. (2013-present)
Demo: <http://nlp.ldeo.columbia.edu/madamira>.

RESEARCH PROTOTYPE TECHNOLOGY

1. **Diab, Mona** and Pradeep Dasigi. (2011) **CODACT** A Tool for Arabic Dialect Orthographic Conventionalization. <https://nlp.ccls.columbia.edu/CODACT>
2. Benajiba, Yassine and **Mona Diab**. (2010) **NIDA** Arabic Named Entity Recognition. Current Version 2.0. <https://nlp.ccls.columbia.edu/NIDA>
3. **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>
4. 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>
5. **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>
6. 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>

PROFESSIONAL ACTIVITIES

ACADEMIC SERVICES WITHIN GW

CS Department Level

- [Sum.2016 -2020] **Member**, Dept. Promotion & Tenure Committee for Prof. Tim Wood
- [Spring 2014] Established in collaboration with other CS faculty *Undergraduate track on Data Science*
- [Fall.2014- present] **Member**, Dept. Committee Grad. Admissions & Support (GAS)
- [Fall.2013 –Fall.2015] **Member**, Dept. Committee on Awards
- [Fall.2014-Spring.2015] **Co-chair**, Big Data Chair Search Committee
- [Fall.2014- Spring.2015] **Member**, Dept. Committee on Budget
- [Fall.2014- Spring.2015] **Member**, Dept. Committee on Student Relations
- [Fall.2013 – Spring.2014] **Member**, Dept. Committee on Research
- [Spring.2013 – Spring.2014] **Member**, Dept. Committee on Graduate Student Admission
- [Spring.2013] **Member**, Dept. Committee on Curriculum
- [Spring.2013] **Member**, Dept. Committee on Faculty Retreat
- [Spring.2013] **Member**, Dept. Committee on Target of Opportunity Hires

SEAS Level

- [Spring.2016] **Participated with feedback on the creation of the MSc in Data Analytics** (with Bhagi Narahari, Claire Monteleoni, Tim Wood, from CS and Faculty from EMSE and Business School). SEAS & Business School
- [Fall.2014 – present] **Member**, Steering Committee IBM collaboration and Point Person for any work or education related to Watson
- [Spring.2014 – Fall.2014] **Elected Senator**, representing SEAS on Faculty Senate
- [Fall.2013 – Spring.2014] **Co-chair**, Big Data Initiative Task Force

GW Level

- [Fall.2015-present] **Member, Vice President of Research Advisory Board**

ACADEMIC SERVICES WITHIN COLUMBIA UNIVERSITY

- [Fall.2005–Fall.2012] **Member** of Columbia Senate Committee for Research Officers,
 - 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.
- [Fall.2005–Fall.2012] **Representing** Columbia University Bi-annual National Council on Women in Information Technology. (NCWIT) meetings
- [Fall.2011–Fall.2012] **Member** of the Columbia Senate Committee on the External Affairs and Public Relations
- [Fall.2006–Fall.2012] **Member** of CCLS Strategy Committee
- [Spring.2010–Fall.2012] **Member** of the Columbia Senate Commission on the Status of Women
- [Fall.2009–Fall.2012] **Member** of SEAS Dean's Committee for Research Scientists
- [Spring.2010–Fall.2012] **Organizer** of CCLS-Yahoo Distinguished Lecture Series
- [Spring.2008–Fall.2009] **Member** of Columbia Senate committee on University Rules of Conduct
- [Spring.2008–Fall.2009] **Member** of Columbia Senate committee on Alumni Relations
- [Fall.2007–Spr.2008] **Organizer** of CCLS Colloquium Series
- [Fall.2007–Spring.2009] **Represented** CCLS in the SEAS Executive Council Meetings
- [Fall.2005–Spring.2006] **Member** of Columbia Senate Committee on Physical Development
- [Spring.2004–Fall.2004] **Organizer** of Stanford Computational Linguistics Colloquium Series

SERVICES WITHIN SCIENTIFIC COMMUNITY

GRANT PROPOSAL REVIEWER

- [03.2022] NSF RI, IIS Panel
- [03.2021] NSF RI, IIS Panel
- [03.2020] NSF RI, IIS Panel
- [03.2019F] NSF RI, IIS Panel
- [12.2015] NSF RI CI Panel
- [01.2012-12.2015] KACST Research Fund
- [03.2013] NSF Semantics Panel
- [02.2012] NSF OISE Program
- [03.2010-09.2012] Qatar National Research Fund
- [04.2011] NSF Linguistics Program
- [01.2010] NSF SoCS Program Panel
- [05.2008] NSF HSD Awards Panel
- [10.2010] NSF Career Awards

EVENT/ CONFERENCE/WORKSHOP ORGANIZATION

[TOTAL: 52]

1. Co-founder of the 1st ACL Arabic NLP Conference (WANLP 2023)
2. Special Theme Chair, Fairness, AACL 2022
3. Diversity & Inclusion Chair, ACL 2022
4. Diversity & Inclusion Chair, WeCNLP 2021
5. Co-Chair, 5th Workshop on Computational Approaches to Linguistic Code switching, NAACL 2021
6. Senior Area Chair, Lexical Semantics, ACL 2021
7. Co-Chair, 4th Workshop on Computational Approaches to Linguistic Code switching, LREC 2020
8. Diversity & Inclusion Co-Chair, WeCNLP 2020
9. Senior Area Chair, Sentence Semantics, ACL 2020
10. Senior Area Chair, Sentence Semantics, ACL 2020
11. PC Chair WeCNLP 2019
12. Senior Area Chair, Lexical Semantics, EMNLP 2019
13. Senior Area Chair, Sentence Level Semantics, ACL 2019
14. Area Chair, Social Media, Computational Social Science, and Sentiment/Opinion Analysis, EMNLP 2018
15. Area Chair, Semantics, NAACL 2018
16. Co-Chair, 3rd Workshop on Computational Approaches to Linguistic Code switching, ACL 2018
17. Area Chair, NLP, Amazon Machine Learning Conference (AMLC), Amazon 2018
18. Workshop chair, Shareable NLP technologies within Amazon, AMLC, 2018
19. Area co-chair Semantics, NAACL 2017
20. Area Co-Chair Multilingual Processing, ACL 2017
21. Faculty organizer and Chair Mid Atlantic Student Colloquium for Speech & Language Learning, 6th MASC-SLL 2017
22. Co-Organizer of PILOT Sentiment & Emotion & Cognitive State Evaluation for DARPA LORELEI
23. PC Co-Chair 2nd Workshop on Arabic Natural Language Processing (WANLP), EACL 2017
24. Area Co-Chair MT and Multilingual Processing, EMNLP 2017
25. PC Co-Chair MT and Multilingual Processing, EACL 2017
26. Co-Chair SemEval 17 Task 1, Semantic Textual Similarity
27. Co-Chair 2nd Workshop on Computational Approaches to Linguistic Code Switching (CALCS), Proceedings of EMNLP 2016, Austin TX, USA
28. Co-Chair Workshop on Machine Translation for Semitic Languages (SeMAT), AMTA 2016, Austin TX, USA
29. Co-Chair Semeval 16 Task 1, Semantic Textual Similarity, NAACL 16, San Diego CA, USA.
30. Chair Nominating committee for NAACL 2016-2018 Board Membership
31. Returning Officer for ACL SIGLEX Board Elections.
32. Co-Chair Semeval 15 Task 2, Semantic Textual Similarity, NAACL 15, Boulder CO, USA.
33. Organizer, Mutliword Expressions Special Event, CICLING 15, Cairo, Egypt.
34. Co-Chair, Town Hall Meeting for Big Data efforts at GW, Washington DC, USA
35. Co-Chair, First Workshop on Computational Approaches to Linguistic Code Switching, Empirical Methods For Natural Language Processing (EMNLP 2014), Doha, Qatar.
36. Publicity Co-Chair, Empirical Methods For Natural Language Processing (EMNLP 2014), Doha, Qatar.
37. Co-Chair Semeval 14 Task 10, Multilingual Semantic Textual Similarity, International Conference of Computational Linguistics (COLING 2014), Dublin, Ireland.
38. General Chair, *SEM 2nd Joint Semantics Conference, 2013

39. *Chair, DARPA Funded Unified Semantic Framework for Textual Similarity Workshop, 2012*
40. *Area Co-Chair, Multilinguality, ACL 2012*
41. *Program Co-Chair of 1st International Joint Conference on Semantics *SEM 12*
42. *Workshops Co-Chair, NAACL 12*
43. *Co-Organizer SEMEVAL 2012 Task on Multilayer Semantic Annotation and STS*
44. *Chair/Co-Organizer of 2nd NSF Funded Workshop on Linguistic Code Switching, 2011*
45. *Chair/Co-Organizer of NSF Funded Workshop on Linguistic Code Switching, 2010*
46. *Area Chair, Semantics and Machine Learning, NAACL 2010*
47. *Area Chair, Multilinguality, EMNLP 2009*
48. *Co-Chair COLING-HLT 2008 Workshop on TextGraphs*
49. *Co-Chair LREC Workshop on Arabic and Local Languages, LREC 2008*
50. *Chair/Co-Organizer of NSF Funded Workshop on Linguistics Code Switching, 2008*
51. *Organizer of SEMEVAL Task on Arabic Semantic Labeling, SemEval 2007*
52. *Co-Chair ACL Workshop on Computational Approaches to Semitic Languages, ACL 2005*

PROGRAM COMMITTEE MEMBER

Journals

Transactions on Computational Linguistics; Journal of AI Research; 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; CACLING, 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)