

Nizar Habash

850 Interchurch, MC 7717
475 Riverside Drive, New York, NY 10115
Office:(212) 870-1289, Cell:(646) 388-2315, Fax:(212) 870-1285
<http://www.NizarHabash.com>
Nizar@NizarHabash.com

Education

Ph.D. Computer Science, University of Maryland College Park, 2003.
M.S. Computer Science, University of Maryland College Park, 2000.
B.S. Computer Engineering, *summa cum laude*, Old Dominion University, 1997.
B.A. Linguistics and Languages, *summa cum laude*, Old Dominion University, 1997.

Employment

Feb. 2009 to present. **Columbia University, Center for Computational Learning Systems, Research Scientist.**
Feb. 2009 to present. **Columbia University, Computer Science Department, Adjunct Associate Professor.**
Jul. 2005 to Jan. 2009. Columbia University, Center for Computational Learning Systems, Associate Research Scientist.
Jul. 2004 to Jun. 2005. Columbia University, Post-doctoral Researcher.
Jul. 2003 to Jun. 2004. University of Maryland College Park, Post-doctoral Researcher.
Jan. 1998 to May 2003. University of Maryland College Park, Graduate Research Assistant.
Jan. 1999 to Jan. 2000. Nuun Labs, Inc., Chief Technical Officer.
Jun. 1998 to Aug. 1998. Cycorp, Inc., Intern Ontological Engineer.
Jan. 1996 to Aug. 1996. Old Dominion University, Undergraduate Research Assistant.

Research Experience

Research Areas

- My main research is natural language processing / computational linguistics.
- My specialty sub-areas are Arabic natural language processing and machine translation.
- Together with Owen Rambow and Mona Diab, I co-founded the Columbia Arabic Dialect Modeling (CADIM) group in 2005.

Past and Current Research Support

- [1] "Preserving Morphological Richness in Poor-Pivot-based Statistical Machine Translation." Nizar Habash (PI). Google Research Award. 9/2012-8/2013 (\$84,211).
- [2] "Automatic Correction of Standard Arabic Text: Resource and System Development." Nizar Habash (PI) and Behrang Mohit. Qatar National Research Fund Award. Approved. \$327,333 per year for three years starting in 2012.
- [3] "Morphology Modeling for LORELEI." Owen Rambow (PI) and Nizar Habash. IBM Subcontract. IARPA BABEL Program. Approved. \$307,500 per year for four years starting in 2012.

- [4] "Arabic Dialect Tool Creation." Nizar Habash (PI), Mona Diab and Owen Rambow. BBN Subcontract. DARPA BOLT Program. \$1,047,619 per year for five years. Year 1 (10/2011-9/2012) approved.
- [5] "Arabic Dialect Morphological Analysis." Nizar Habash (PI), Mona Diab and Owen Rambow. LDC Subcontract. DARPA BOLT Program. Year 1 (\$88,142) and Year 2 option (\$91,443). Year 1 (10/2011-9/2012) approved.
- [6] "Language Modeling for PLATO." Nizar Habash (PI), Mona Diab and Owen Rambow. BBN Subcontract. DARPA MADCAT Program. Five years: P1 10/2007-4/2009 (\$192,500); P2 5/2009-5/2010 (\$229,934); P3 4/2010-9/2011 (255,582); P4 9/2011-5/2012 (120,000); and P5 optional (\$153,661).
- [7] "Arabic Generation in Statistical Machine Translation." Nizar Habash (PI). Google Research Award. Year1: 9/2009-8/2010 (\$60,000) and Year 2: 9/2010-8/2011 (\$75,000).
- [8] "Arabic Dialect Tools and Resources for COLABA." Mona Diab (PI), Nizar Habash and Owen Rambow. Axiom Subcontract. TSWG COLABA Program. P1 10/2008-3/2010 (\$442,709); P2 4/2010-3/2011(\$257,618)
- [9] "Arabic Dialect Translation for Agile." Nizar Habash (PI). IBM Subcontract. DARPA GALE Program. GALE-Y5 6/2010-6/2011 (\$112,437).
- [10] "Arabic NLP/MT for Rosetta." Owen Rambow (PI), Mona Diab and Nizar Habash. IBM Subcontract. DARPA GALE Program. GALE-Y3 4/2008-6/2009 (\$385,116); GALE-Y4 6/2009-6/2010 (\$509,586); GALE-Y5 7/2010-6/2011 (\$300,000).
- [11] "Columbia Arabic Tree Bank." Nizar Habash (PI). IBM Subcontract. DARPA GALE Program. GALE-Y3 4/2008-8/2008 (\$200,000).
- [12] "Arabic NLP/MT for NIGHTINGALE." Owen Rambow (PI), Mona Diab and Nizar Habash. SRI Subcontract. DARPA GALE Program. GALE-Y1: 9/2005-8/2006 (\$449,214) and GALE-Y2: 9/2006-8/2007 (\$493,262).
- [13] "CINDOR: Conceptual Interlingua Document Retrieval." Owen Rambow (PI) and Nizar Habash (Post-doc). Textwise, Inc. 2/2005-5/2006 (\$38,534).
- [14] "Parsing Arabic Dialects". Owen Rambow (team leader), Mona Diab (senior researcher) and Nizar Habash (senior researcher). An NSF-funded Johns Hopkins Summer Workshop, 2005 (\$15,750).
- [15] "Arabic Dialect Modeling for Speech and Natural Language Processing." Owen Rambow (PI) and Nizar Habash (Post-doc). NSF. 7/2004-8/2006 (\$200,000).

Research Advising/Supervision

- I am currently directing the following doctoral students:

- [1] Ahmed El Kholy, Ph.D. student, Computer Science Department, Columbia University. 2008-present.
- [2] Sarah Alkuhlani, Ph.D. student, Computer Science Department, Columbia University. 2009-present.
- [3] Wael Salloum, Ph.D. student, Computer Science Department, Columbia University. 2010-present.

- I have served as the primary or partial advisor/supervisor of the following researchers:

Junior Colleagues

- [4] Ryan Roth, Staff Associate, Columbia University Center for Computational Learning Systems (with Mona Diab and Owen Rambow). 2010-present.

Post-doctoral

- [5] Abdelati Hawwari, Post-doctoral researcher, Columbia University Center for Computational Learning Systems (with Mona Diab and Owen Rambow). 2011-present.
- [6] Marine Carpuat, Post-doctoral researcher, Columbia University Center for Computational Learning Systems (with Mona Diab and Owen Rambow). 2010-2011.
- [7] Yuval Marton, Post-doctoral researcher, Columbia University Center for Computational Learning Systems. (Primary research advisor). 2009-2011.
- [8] Ryan Roth, Post-doctoral researcher, Columbia University Center for Computational Learning Systems (with Mona Diab and Owen Rambow). 2008-2009.

Doctoral

- [9] Kristen Parton, Ph.D. student, Computer Science Department, Columbia University. (Advisor: Kathy McKeown). 2010-present.
- [10] Kais Dukes, Ph.D. student, University of Leeds. (Advisor: Eric Atwell). 2009-present.
- [11] Mohamed El-Tantawy, Ph.D. student, Computer Science Department, Columbia University. (Advisor: Owen Rambow). 2008-2011.
- [12] Fadi Biadsy, Ph.D. student, Computer Science Department, Columbia University. (Advisor: Julia Hirschberg). 2008-2011.
- [13] Josep M Crego. Ph.D. student, Computer Science Department, Universitat Politècnica de Catalunya. (Advisor: José Mariño). 2007-2009
- [14] Jakob Elming, Ph.D. student, Computer Science Department, Copenhagen Business School. (Advisor: Daniel Hardt). 2006-2009.
- [15] Mahmoud Ghoneim. Ph.D. student, Computer Science Department, Cairo University. (Columbia Advisor: Mona Diab). 2006-2009.
- [16] Necip Fazil Ayan, Ph.D. student. Computer Science Department, University of Maryland College Park. (Advisor: Bonnie Dorr). 2003-2005.

Masters

- [17] Amit Abbi. M.S. student, Computer Science Department, Columbia University (with Owen Rambow and Mona Diab). 2010-2011.
- [18] Reshef Shilon, M.S. student, Tel Aviv University. (Advisor: Shuly Wintner). 2010-2011.
- [19] Jun Hu, M.S. student, Computer Science Department, Columbia University. (Primary research advisor). 2008-2009.
- [20] Ryan Roth, M.S. student, Computer Science Department, Columbia University (with Owen Rambow). 2007.
- [21] Abdelhalim Abbas, M.S. student, Computer Science Department, Columbia University. (Primary research advisor). 2006.
- [22] Farhad Peydaye Saheli, M.S. student, Computer Science Department, Columbia University (with Owen Rambow). 2006.
- [23] Fadi Biadsy, M.S. student, Computer Science Department, Ben-Gurion University. (Advisor: Jihad El-Sana). 2005-2006.
- [24] Matthew Chu, M.S. student, Computer Science Department, Columbia University (with Owen Rambow). 2005.

Undergraduate

- [25] Jacob Andreas, undergraduate student, Computer Science Department, Columbia University (with Owen Rambow). 2010-2011.
- [26] Morgan Rhodes, undergraduate student, Computer Science Department, Columbia University. (Primary research advisor). 2006-2007.

High School

- [27] Hayden Metsky, High school student, Millburn High School, NJ. (Primary research advisor). 2007-2009. Under my mentorship, this student was the regional winner in the Siemens Competition in Math, Science and Technology (top 6 out of 1893) and a finalist in the Intel Science Talent Search (top 40 out of 1600), 2009.

Teaching Experience

Classes Taught

- [1] Instructor of graduate-level course. E6998 section 003: Machine Translation. Department of Computer Science, Columbia University, Fall 2011. 9 students.
[Syllabus: <https://sites.google.com/site/comse6998machinetranslation/>]
- [2] Instructor of graduate-level course. E6998 section 003: Machine Translation. Department of Computer Science, Columbia University, Fall 2010. 21 students.

The following is a summary of the student evaluation for the above course (n=10). Score are between 1 and 5 (5 is Excellent); [Mean/Median/Std.Dev.]:

1: Instructor: Organization and Preparation	[4.50/4.5/0.53]
2: Instructor: Classroom Delivery	[4.70/5.0/0.48]
3: Instructor: Approachability	[5.00/5.0/0.00]
4: Instructor: Overall Quality	[4.80/5.0/0.42]
5: Course: Amount Learned	[4.90/5.0/0.32]
6: Course: Appropriateness of Workload	[4.70/5.0/0.48]
7: Course: Fairness of Grading Process	[4.60/5.0/0.70]
8: Course: Overall Quality	[4.80/5.0/0.42]

- [3] Co-instructor with Professor Bonnie Dorr. CMSC 723/LING 645: Introduction to Computational Linguistics University of Maryland College Park, Spring 2004. 35 students.

Teaching Materials

I wrote a textbook, “[Introduction to Arabic Natural Language Processing](#)” as part of the [Synthesis Lectures on Human Language Technology](#) series published by Morgan & Claypool Publishers. The book is currently being translated to Arabic by Dr. Hend Al-Khalifa of King Saud University.

I created two popular tutorials, one on [Arabic Natural Language Processing](#) and one (with my colleague Mona Diab) on [Arabic Dialect Processing](#).

Teaching Service

I am the Ph.D. student coordinator for the Center for Computational Learning Systems (CCLS).

Guest Lectures

- [1] **Habash, Nizar.** [Introduction to Arabic Natural Language Processing](#), CS 4705: Introduction to Natural Language Processing, Columbia University, Fall 2011.
- [2] **Habash, Nizar.** [Machine Translation: Challenges, Approaches and Evaluation](#), CS 4705: Introduction to Natural Language Processing, Columbia University, Fall 2010, Fall 2008-2004.
- [3] **Habash, Nizar.** [Machine Translation: Challenges and Approaches](#), CMSC 421: Introduction to Artificial Intelligence, University of Maryland College Park, Fall 2003.
- [4] **Habash, Nizar.** [Generation Heavy Hybrid Machine Translation](#), CMSC 723/LING 645: Introduction to Computational Linguistics, University of Maryland College Park, Spring 2003.
- [5] **Habash, Nizar.** [Computers and Writing Systems](#), HONR 279I: History of the Alphabets: Languages and their Scripts, University of Maryland College Park, Spring 2003.
- [6] **Habash, Nizar.** [Natural Language Generation](#), CMSC 828: Advanced Natural Language Processing: Theory and Practice, University of Maryland College Park, Spring 2002.

Conference Tutorials

See also *Invited Tutorials* under *Oral Presentations*

- [1] **Habash, Nizar.** [Introduction to Arabic Natural Language Processing](#).
 - A two-hour tutorial, 80 attendees. In MEDAR: The 2nd International Conference on Arabic Language Resources and Tools, Cairo, Egypt, 2009.
 - A three-hour tutorial, 20 attendees. In the International Conference on Language Resources and Evaluation (LREC), Genoa, Italy, 2006.
 - A three-hour tutorial, 35 attendees. In the Association for Computational Linguistics Conference (ACL'05), Ann Arbor, Michigan, 2005.
 - A three-hour tutorial, 17 attendees. In the Association for Machine Translation in the Americas conference (AMTA'04), Washington DC, 2004.

- [2] Diab, Mona and **Nizar Habash**. Arabic Dialect Processing.
- A two-hour tutorial, 80 attendees. In MEDAR: The 2nd International Conference on Arabic Language Resources and Tools, Cairo, Egypt, 2009.
 - A three-hour tutorial, 30 attendees. In the International Conference on Language Resources and Evaluation (LREC), Marrakech, Morocco, 2008.
 - A three-hour tutorial, 15 attendees. In the Conference of the North American Association for Computational Linguistics (NAACL'07), Rochester, NY, 2007.
 - A three-hour tutorial, 8 attendees. In the Association for Machine Translation in the Americas (AMTA'06), Boston, MA, 2006.
- [3] **Habash, Nizar**. Arabic Natural Language Processing for Machine Translation. A three-hour tutorial, 12 attendees. In the meeting of the Association for Machine Translation in the Americas conference (AMTA'08), Waikiki, Hawai'i, 2008.

Publications

Journal Articles

- [1] Ahmed El Kholy and **Nizar Habash**. "Orthographic and Morphological Processing for English-Arabic Statistical Machine Translation". In Machine Translation Journal Special Issue on Arabic Machine Translation, 2011.
- [2] Marine Carpuat, Yuval Marton and **Nizar Habash**. "Reordering Post-verbal Subjects for Arabic-to-English Statistical Machine Translation". In Machine Translation Journal Special Issue on Arabic Machine Translation, 2011.
- [3] Shilon, Reshef, **Nizar Habash**, Alon Lavie, Shuly Wintner. "Machine Translation between Hebrew and Arabic". In Machine Translation Journal Special Issue on Arabic Machine Translation, 2011.
- [4] Kais Dukes, Eric Atwell and **Nizar Habash**. "Supervised Collaboration for Syntactic Annotation of Quranic Arabic", Language Resource and Evaluation Journal, 2011.
- [5] Dorr, B., R. Passonneau, D. Farwell, R. Green, **N. Habash**, S. Helmreich, E. Hovy, L. Levin, K. Miller, T. Mitamura, O. Rambow, F. Reeder, and A. Siddharthan. "Interlingual Annotation of Parallel Text Corpora: A New Framework for Annotation and Evaluation," Natural Language Engineering, Cambridge University Press, 2010.
- [6] **Habash, Nizar**, Bonnie J. Dorr and Chirstof Monz. Symbolic-to-Statistical Hybridization: Extending Generation-Heavy Machine Translation. Machine Translation Journal, pages 23--63, 2009.
- [7] Dorr, Bonnie J., Necip Fazil Ayan, **Nizar Habash**, Nitin Madnani, and Rebecca Hwa. Rapid Porting of DUSTer to Hindi. ACM Transactions on Asian Language Information Processing, 2:3, 2003.
- [8] **Habash, Nizar**, Bonnie Dorr, and David Traum. Hybrid Natural Language Generation from Lexical Conceptual Structures. MT Journal volume 18 (2): 81-128, 2003.
- [9] **Habash, Nizar**. "Nuun: A System for Developing Platform and Browser Independent Arabic Web Applications." (In Arabic) Arab Journal of Science, 33, 1999.

Books

- [1] **Habash, Nizar**. Introduction to Arabic Natural Language Processing, Synthesis Lectures on Human Language Technologies, Graeme Hirst, editor. Morgan & Claypool Publishers. 187 pages, 2010.

Book Chapters

- [1] **Nizar Habash** and Fatiha Sadat. "Chapter 5: Arabic Preprocessing for Statistical Machine Translation: Schemes, Techniques and Combinations." In Souidi, A., Neumann, G., Farghaly, A., Zbib, R. eds., "Challenges for Arabic MT", John Benjamins NLP series, 2012.
- [2] Diab, Mona, **Nizar Habash**, Owen Rambow, Ryan Roth. "CADIM Arabic Tools: Morphological Analysis, Disambiguation and Generation, Tokenization, Diacritization, Lemmatization, POS Tagging

- and Base Phrase Chunking.” In Chapter 1 on Data Acquisition and Linguistic Resources. In the Handbook of Natural Language Processing and Machine Translation, Joseph Olive, Caitlin Christianson and John McCary, eds. Springer, 2011.
- [3] Maamouri, Mohamed, Ann Bies, Seth Kulick, **Habash, Nizar**, Reem Faraj and Ryan Roth. “Arabic Treebanking.” In Chapter 1 on Data Acquisition and Linguistic Resources. In the Handbook of Natural Language Processing and Machine Translation, Joseph Olive, Caitlin Christianson and John McCary, eds. Springer, 2011.
- [4] De Gispert, Adrià, William Byrne, Jinxi Xu, Rabih Zbib, John Makhoul, Achraf Chalabi, Hazem Nader, **Nizar Habash** and Fatiha Sadat. “Preprocessing Arabic for Arabic-English Statistical Machine Translation.” In Chapter 2 on Machine Translation from Text. In the Handbook of Natural Language Processing and Machine Translation, Joseph Olive, Caitlin Christianson and John McCary, eds. Springer, 2011.
- [5] **Habash, Nizar**, “Introduction.” In Chapter 2 on Machine Translation from Text. In the Handbook of Natural Language Processing and Machine Translation, Joseph Olive, Caitlin Christianson and John McCary, eds. Springer, 2011.
- [6] **Habash, Nizar**. “Using Surface Span in Syntax-aware MT Evaluation (SEPIA).” In Chapter 5 on Machine Translation Evaluation and Optimization. In the Handbook of Natural Language Processing and Machine Translation, Joseph Olive, Caitlin Christianson and John McCary, eds. Springer, 2011.
- [7] Farwell, D., B. Dorr, **N. Habash**, S. Helmreich, E. Hovy, R. Green, L. Levin, K. Miller, T. Mitamura, O. Rambow, F. Reeder, and A. Siddharthan. “Chapter 10: Interlingual Annotation of Multilingual Text Corpora and FrameNet.” In: H.C. Boas (ed.), /Multilingual FrameNets in Computational Lexicography: Methods and Applications/, 287-318. Berlin: Mouton de Gruyter, 2009.
- [8] Elming, Jakob, **Nizar Habash** and Josep Crego. “Combination of Statistical Word Alignments Based on Multiple Preprocessing Schemes.” Book chapter in “Learning for Machine Translation.” Editors Cyril Goutte, Nicola Cancedda, Marc Dymetman, and George Foster. MIT Press, 2008.
- [9] **Habash, Nizar**. “Arabic Morphological Representations for Machine Translation.” Book Chapter. In Arabic Computational Morphology: Knowledge-based and Empirical Methods. Editors Antal van den Bosch and Abdelhadi Soufi. 2007.
- [10] **Habash, Nizar**, Abdelhadi Soufi, and Tim Buckwalter. “On Arabic Transliteration.” Book Chapter. In Arabic Computational Morphology: Knowledge-based and Empirical Methods. Editors Antal van den Bosch and Abdelhadi Soufi. 2007.

Conference Proceedings

- [1] Ahmed El Kholy and **Nizar Habash**. Automatic error analysis for morphologically rich languages. In Proceedings of the Thirteenth Machine Translation Summit (MT Summit XIII), pp.225-232, Xiamen, China, 2011.
- [2] Dukes, Kais and **Nizar Habash**. One-Step Statistical Parsing of Hybrid Dependency-Constituency Syntactic Representations. In Proceedings of the International Conference on Parsing Technologies. Dublin, 2011.
- [3] Alkuhlani, Sarah and **Nizar Habash**. A Corpus for Modeling Morpho-Syntactic Agreement in Arabic: Gender, Number and Rationality. In Proceedings of Conference of the Association for Computational Linguistics (ACL), Short Paper, Portland, OR, 2011.
- [4] **Habash, Nizar** and Ryan Roth. Using Deep Morphology to Improve Automatic Error Detection in Arabic Handwriting Recognition. In Proceedings of the Conference of the Association for Computational Linguistics (ACL), Portland, OR, 2011.
- [5] Marton, Yuval, **Nizar Habash** and Owen Rambow. Improving Arabic Dependency Parsing with Lexical and Inflectional Surface and Functional Features. In Proceedings of the Conference of the Association for Computational Linguistics (ACL), Portland, OR, 2011.
- [6] Carpuat, Marine, Yuval Marton and **Nizar Habash**. Improving Arabic-to-English Statistical Machine Translation by Reordering Post-verbal Subjects for Alignment. In Proceedings of the Conference of the Association for Computational Linguistics (ACL). Uppsala, Sweden, 2010.
- [7] Carpuat, Marine, Yuval Marton and **Nizar Habash**. Reordering Matrix Post-verbal Subjects for Arabic-to-English SMT. In Proceedings of Traitement Automatique du Langage Naturel (TALN-10). Montréal, Canada, 2010. [Best Paper Award]

- [8] Kholy, Ahmed El and **Nizar Habash**. Orthographic and Morphological Processing for English-Arabic Statistical Machine Translation. In Proceedings of Traitement Automatique du Langage Naturel (TALN-10). Montréal, Canada, 2010.
- [9] Mohamed Altantawy, **Nizar Habash**, Owen Rambow, and Ibrahim Saleh. Morphological Analysis and Generation of Arabic Nouns: A Morphemic Functional Approach. In Proceedings of the Language Resources and Evaluation Conference (LREC), Malta, 2010.
- [10] Kais Dukes and **Nizar Habash**. Morphological Annotation of Quranic Arabic. In Proceedings of the Language Resources and Evaluation Conference (LREC), Malta, 2010.
- [11] **Habash, Nizar** and Ryan Roth. CATiB: The Columbia Arabic Treebank. In the Proceedings of the Conference of the Association for Computational Linguistics (ACL'09). Suntec, Singapore, 2009.
- [12] Biadisy, Fadi, **Nizar Habash** and Julia Hirschberg. Improving the Arabic Pronunciation Dictionary for Phone and Word Recognition with Linguistically-Based Pronunciation Rules. In Proceedings of the North American Association for Computational Linguistics (NAACL), Boulder, 2009.
- [13] **Habash, Nizar**, Owen Rambow and Ryan Roth. MADA+TOKAN: A Toolkit for Arabic Tokenization, Diacritization, Morphological Disambiguation, POS Tagging, Stemming and Lemmatization. In Proceedings of the 2nd International Conference on Arabic Language Resources and Tools (MEDAR), Cairo, Egypt, 2009.
- [14] **Habash, Nizar**. REMOOV: A Tool for Online Handling of Out-of-Vocabulary Words in Machine Translation. In Proceedings of the 2nd International Conference on Arabic Language Resources and Tools (MEDAR), Cairo, Egypt, 2009.
- [15] **Habash, Nizar**, Reem Faraj and Ryan Roth. Syntactic Annotation in the Columbia Arabic Treebank. In Proceedings of the 2nd International Conference on Arabic Language Resources and Tools (MEDAR), Cairo, Egypt, 2009.
- [16] **Habash, Nizar** and Hayden Metsky. Automatic Learning of Morphological Variations for Handling Out-of-Vocabulary Terms in Urdu-English Machine Translation. In Proceedings of the Association for Machine Translation in the Americas (AMTA-2008), Waikiki, Hawai'i, 2008.
- [17] **Habash, Nizar**. Four Techniques for Online Handling of Out-of-Vocabulary Words in Arabic-English Statistical Machine Translation. In Proceedings of Association for Computational Linguistics (ACL), Columbus, Ohio. 2008.
- [18] Roth, Ryan, Owen Rambow, **Nizar Habash**, Mona Diab, and Cynthia Rudin. Arabic Morphological Tagging, Diacritization, and Lemmatization Using Lexeme Models and Feature Ranking. In Proceedings of Association for Computational Linguistics (ACL), Columbus, Ohio. 2008.
- [19] **Habash, Nizar** and Ryan Roth. Identification of Naturally Occurring Numerical Expressions in Arabic. In Proceedings of the Language Resources and Evaluation Conference (LREC), Marrakech, Morocco, 2008.
- [20] Farber, Benjamin, Dayne Freitag, **Nizar Habash** and Owen Rambow. Improving NER in Arabic Using a Morphological Tagger. In Proceedings of the Language Resources and Evaluation Conference (LREC), Marrakech, Morocco, 2008.
- [21] **Habash, Nizar**. Syntactic Preprocessing for Statistical Machine Translation, In Proceedings of the Machine Translation Summit (MT-Summit), Copenhagen, Denmark, 2007.
- [22] Diab, Mona, Mahmoud Ghoneim and **Nizar Habash**. Arabic Diacritization in the Context of Statistical Machine Translation, In Proceedings of the Machine Translation Summit (MT-Summit), Copenhagen, Denmark, 2007.
- [23] Kirchhoff, Katrin, Owen Rambow, **Nizar Habash**, Mona Diab. Semi-Automatic Error Analysis for Large-Scale Statistical Machine Translation Systems, In Proceedings of the Machine Translation Summit (MT-Summit), Copenhagen, Denmark, 2007.
- [24] **Habash, Nizar**, Ryan Gabbard, Owen Rambow, Seth Kulick and Mitch Marcus. Determining Case in Arabic: Learning Complex Linguistic Behavior Requires Complex Linguistic Features, In Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP), Prague, Czech Republic, 2007.
- [25] **Habash, Nizar** and Owen Rambow. Arabic Diacritization through Full Morphological Tagging, In Proceedings of the North American chapter of the Association for Computational Linguistics (NAACL), Rochester, New York, 2007.
- [26] Elming, Jakob and **Nizar Habash**. Combination of Statistical Word Alignments Based on Multiple Preprocessing Schemes, In Proceedings of the North American chapter of the Association for Computational Linguistics (NAACL), Rochester, New York, 2007.

- [27] **Habash, Nizar** and Owen Rambow. Morphophonemic and Orthographic Rules in a Multi- Dialectal Morphological Analyzer and Generator for Arabic Verbs, International Symposium on Computer and Arabic Language (ISCAL), Riyadh, Saudi Arabia, 2007.
- [28] **Habash, Nizar**, Bonnie Dorr and Christof Monz. Challenges in Building an Arabic Generation-Heavy Machine Translation System and Extending it with Statistical Components. In Proceedings of the Association for Machine Translation in the Americas (AMTA-2006), Boston, MA, 2006.
- [29] **Habash, Nizar** and Owen Rambow. MAGEAD: a morphological analyzer and generator for the Arabic dialects. In Proceedings of COLING-ACL, Sydney, Australia, 2006.
- [30] Sadat, Fatiha and **Nizar Habash**. Morphological Preprocessing Scheme Combination for Statistical MT. In Proceedings of COLING-ACL, Sydney, Australia, 2006.
- [31] **Habash, Nizar** and Fatiha Sadat. Arabic Preprocessing Schemes for Statistical Machine Translation, In Proceedings of the North American chapter of the Association for Computational Linguistics (NAACL), New York, 2006.
- [32] Chiang, David, Mona Diab, **Nizar Habash**, Owen Rambow, and Safi Shareef. Arabic Dialect Parsing. In Proceedings of the European chapter of the Association of Computational Linguistics (EACL). 2006.
- [33] **Habash, Nizar**, Clinton Mah, Randy Calistri-Yeh, Sabiha Imran and Paraic Sheridan. The Design and Validation of an Arabic WordNet for Information Retrieval. In Proceedings of the International Conference on Language Resources and Evaluation (LREC). 2006.
- [34] Rambow, Owen, Bonnie Dorr, David Farwell, Rebecca Green, **Nizar Habash**, Stephen Helmreich, Eduard Hovy, Lori Levin, Carnegie Keith J. Miller, Teruko Mitamura, Florence Reeder, Advait Siddharthan. Parallel Syntactic Annotation of Multiple Languages. In Proceedings of the International Conference on Language Resources and Evaluation (LREC). 2006.
- [35] Passonneau, Rebecca, **Nizar Habash** and Owen Rambow. Inter-annotator Agreement on a Multilingual Semantic Annotation Task. In Proceedings of the International Conference on Language Resources and Evaluation (LREC). 2006.
- [36] Maamouri, Mohamed, Ann Bies, Tim Buckwalter, Mona Diab, **Nizar Habash**, Owen Rambow, Dalila Tabessi. Developing and Using a Pilot Dialectal Arabic Treebank. In Proceedings of the International Conference on Language Resources and Evaluation (LREC). 2006.
- [37] **Habash, Nizar** and Owen Rambow. Arabic tokenization, part-of-speech tagging and morphological disambiguation in one fell swoop. In Proceedings of the Conference of American Association for Computational Linguistics (ACL), 2005.
- [38] **Habash, Nizar**. The Use of a Structural N-gram Language Model in Generation-Heavy Hybrid Machine Translation. In Proceedings of the Third International Conference of Natural Language Generation (INLG-04). Careys Manor, UK, 2004.
- [39] **Habash, Nizar**. Large Scale Lexeme Based Arabic Morphological Generation. In Proceedings of Traitement Automatique du Langage Naturel (TALN-04). Fez, Morocco, 2004.
- [40] **Habash, Nizar** and Owen Rambow. Extracting a Tree Adjoining Grammar from the Penn Arabic Treebank. In Proceedings of Traitement Automatique du Langage Naturel (TALN-04). Fez, Morocco, 2004.
- [41] Ayan, Fazil, Bonnie J. Dorr, and **Nizar Habash**, Application of Alignment to Real-World Data: Combining Linguistic and Statistical Techniques for Adaptable MT. In Proceedings of the 6th Conference of the Association for Machine Translation in the Americas (AMTA-2004), Georgetown University, Washington DC, 2004.
- [42] Reeder, Florence, Bonnie Dorr, David Farwell, **Nizar Habash**, Stephen Helmreich, Eduard Hovy, Lori Levin, Teruko Mitamura, Keith Miller, Owen Rambow, Advait Siddharthan. Interlingual Annotation for MT Development. In Proceedings of the 6th Conference of the Association for Machine Translation in the Americas (AMTA-2004), Georgetown University, Washington DC, 2004.
- [43] **Habash, Nizar**. Matador: A Large Scale Spanish-English GHMT System. In Proceedings of the MT Summit, New Orleans, LA, 2003.
- [44] **Habash, Nizar** and Bonnie Dorr. A Categorical Variation Database for English. In Proceedings of North American Association for Computational Linguistics, Edmonton, Canada, 2003.
- [45] **Habash, Nizar** and Bonnie Dorr. Handling Translation Divergences: Combining Statistical and Symbolic Techniques in Generation-Heavy Machine Translation. In Proceedings of the Conference of the Association for Machine Translation in the Americas (AMTA), Tiburon, CA, 2002.

- [46] Dorr, Bonnie, Lisa Pearl, Rebecca Hwa and **Nizar Habash**. DUSTER: A Method for Unraveling Cross-Language Divergences for Statistical Word-Level Alignment. In Proceedings of the Conference of the Association for Machine Translation in the Americas (AMTA), Tiburon, CA, 2002.
- [47] **Habash, Nizar**. Generation-Heavy Hybrid Machine Translation. In Proceedings of the International Natural Language Generation Conference (NLG). New York, 2002.
- [48] Habash, Nizar and Bonnie Dorr. Large Scale Language Independent Generation Using Thematic Hierarchies. In Proceedings of the MT Summit VIII. Santiago de Compostella, Spain, 2001.
- [49] **Habash, Nizar**. oxyGen: A Language Independent Language Realization Engine. In Proceedings of the Fourth Conference of the Association for Machine Translation in the Americas (AMTA), Cuernavaca, Mexico, 2000.
- [50] **Habash, Nizar**. Nuun: A System for Developing Platform and Browser Independent Arabic Web Applications. In Proceedings of the Arabic Translation and Localization Conference (ATLAS), Tunis, Tunisia, 1999.
- [51] **Habash, Nizar**. Issues in Palestinian Arabic Spelling Standardization. In the North American Conference on Afro-Asiatic Linguistics (NACAL 27), Baltimore, MD, 1999.
- [52] Dorr, Bonnie, **Nizar Habash** and David Traum. A Thematic Hierarchy for Efficient Generation from Lexical-Conceptual Structures. In Proceedings of the Association of Machine Translation in the Americas (AMTA), Longhorne, PA, 1998.

Workshops

- [1] Jacob Andreas, **Nizar Habash**, and Owen Rambow. Fuzzy syntactic reordering for phrase-based statistical machine translation. In Proceedings of the Workshop on Statistical Machine Translation, Edinburgh, Scotland, UK, 2011.
- [2] Yuval Marton, Ahmed El Kholly and **Nizar Habash**. Filtering antonymous, trend-contrasting, and polarity-dissimilar distributional paraphrases for improving statistical machine translation. In Proceedings of the Workshop on Statistical Machine Translation, Edinburgh, Scotland, UK, 2011.
- [3] Salloum, Wael and **Nizar Habash**. Dialectal to Standard Arabic Paraphrasing to Improve Arabic-English Statistical Machine Translation. In proceedings of the Dialects workshop at the Conference for Empirical Methods in Natural Language Processing (EMNLP 2011). Edinburgh, UK, 2011.
- [4] Altantawy, Mohamed, **Nizar Habash** and Owen Rambow. Fast Yet Rich Morphological Analysis. In proceedings of the 9th International Workshop on Finite-State Methods and Natural Language Processing (FSMNLP 2011), Blois, France, 2011.
- [5] Marton, Yuval, **Nizar Habash** and Owen Rambow. Improving Arabic Dependency Parsing with Lexical and Inflectional Morphological Features. In Proceedings of the North American Association for Computational Linguistics (NAACL) Workshop on Statistical Parsing of Morphologically Rich Languages (SPMRL 2010), Los Angeles, 2010.
- [6] Kholly, Ahmed El and **Nizar Habash**. Techniques for Arabic Morphological Detokenization and Orthographic Denormalization. In Proceedings of the Language Resources and Evaluation Conference (LREC) Workshop on Semitic Languages, Malta, 2010.
- [7] Diab, Mona, **Nizar Habash**, Owen Rambow, Mohamed Altantawy and Yassine Benajiba. COLABA: Arabic Dialect Annotation and Processing. In Proceedings of the Language Resources and Evaluation Conference (LREC) Workshop on Semitic Languages, Malta, 2010.
- [8] Shilon, Reshsef, **Nizar Habash**, Alon Lavie and Shuly Wintner. Machine Translation between Hebrew and Arabic: Needs, Challenges and Preliminary Solutions. Student Research Workshop, In Proceedings of the Association of Machine Translation in the Americas (AMTA), 2010.
- [9] Saleh, Ibrahim and **Nizar Habash**. Automatic Extraction of Lemma-based Bilingual Dictionaries for Morphologically Rich Languages. In the Third Workshop on Computational Approaches to Arabic Script-based Languages at the Machine Translation Summit XII Ottawa, Ontario, Canada, 2009.
- [10] **Habash, Nizar** and Jun Hu. Improving Arabic-Chinese Statistical Machine Translation using English as Pivot Language. In Proceedings of the Workshop on Statistical Machine Translation at the meeting of the European Association for Computational Linguistics (EACL), Athens, Greece, 2009.
- [11] Biadisy, Fadi, Julia Hirschberg and **Nizar Habash**. Spoken Arabic Dialect Identification Using Phonotactic Modeling. In Proceedings of the Workshop on Computational Approaches to Semitic Languages at the meeting of the European Association for Computational Linguistics (EACL), Athens, Greece, 2009.

- [12] Elming, Jakob and **Nizar Habash**. Syntactic Reordering for English-Arabic Phrase-Based Machine Translation. In Proceedings of the Workshop on Computational Approaches to Semitic Languages at the meeting of the European Association for Computational Linguistics (EACL), Athens, Greece, 2009.
- [13] Buckwalter, Tim and **Nizar Habash**. Buckwalter Arabic Morphological Analyzer (BAMA) in Arabic Presentation Form. A participating system in the Arab League Educational, Cultural and Scientific Organization (ALECSO) workshop on Arabic Morphological Analyzers, Damascus, Syria, 2009.
- [14] **Habash, Nizar** and Ahmed El Kholy. SEPIA: Surface Span Extension to Syntactic Dependency Precision-based MT Evaluation. In Proceedings of the workshop on Metrics for Machine Translation at the meeting of Association for Machine Translation in the Americas (AMTA), Waikiki, Hawai'i, 2008.
- [15] Crego, Josep M. and **Nizar Habash**. Using Shallow Syntax Information to Improve Word Alignment and Reordering for SMT. In Proceedings of the Statistical Machine Translation Workshop at the Conference of Association for Computational Linguistics (ACL), Columbus, Ohio, 2008.
- [16] **Habash, Nizar**, Owen Rambow, Mona Diab and Reem Faraj. Guidelines for Annotating Arabic Dialect. In Proceedings of Workshop on Arabic and its local languages (LREC), Marrakech, Morocco, 2008.
- [17] Biadisy, Fadi, Jihad El-Sana and **Nizar Habash**. Arabic Online Handwriting Recognition. International Workshop on Handwriting and Optical Character Recognition, Paris, France, 2006.
- [18] **Habash, Nizar**, Owen Rambow and George Kiraz. Morphological Analysis and Generation for Arabic Dialects. In Proceedings of the Workshop on Computational Approaches to Semitic Languages at the Conference of American Association for Computational Linguistics (ACL), 2005.
- [19] Farwell, David, Stephen Helmreich, Bonnie J. Dorr, **Nizar Habash**, Florence Reeder, Keith Miller, Lori Levin, Teruko Mitamura, Eduard Hovy, Owen Rambow, and Advait Siddharthan. Interlingual Annotation of Multilingual Text Corpora. In Proceedings of the North American Chapter of the Association for Computational Linguistics (NAACL) Workshop on Frontiers in Corpus Annotation, Boston, 2004.
- [20] Mitamura, Teruko, Keith J. Miller, Bonnie J. Dorr, David Farwell, **Nizar Habash**, Lori Levin, Stephen Helmreich, Eduard Hovy, Lori Levin, Owen Rambow, Reeder, Florence, and Advait Siddharthan. Semantic Annotation of Multilingual Text Corpora. In Proceedings of the Workshop on Beyond Named Entity Recognition: Semantic Labeling for NLP Tasks, LREC, Portugal, 2004.
- [21] Dorr, Bonnie J., Rebecca Green, Lori Levin, Owen Rambow, David Farwell, **Nizar Habash**, Stephen Helmreich, Eduard Hovy, Keith J. Miller, Teruko Mitamura, Florence Reeder, and Advait Siddharthan. Semantic Annotation and Lexico-Syntactic Paraphrase. In Proceedings of the Workshop on Building Lexical Resources from Semantically Annotated Corpora, LREC, Portugal, 2004.
- [22] Dorr, Bonnie and **Nizar Habash**. Interlingua Approximation: A Generation-Heavy Approach. In Proceedings of the Workshop on Interlingua Reliability, Fifth Conference of the Association for Machine Translation in the Americas, AMTA-2002, Tiburon, CA, 2002.
- [23] Traum, David and **Nizar Habash**. Generation from Lexical Conceptual Structures. Workshop on Applied Interlinguas, ANLP-2000. Seattle, WA.

Technical Reports

- [1] **Habash, Nizar**, Mona Diab, and Owen Rambow. CODA: Conventional Orthography for Dialectal Arabic. Center for Computational Learning Systems technical report #CCLS-11-02. 2011.
- [2] **Habash, Nizar**, Owen Rambow and Ryan Roth. MADA+TOKAN Manual. Center for Computational Learning Systems technical report #CCLS-10-01. 2010.
- [3] Bonnie J. Dorr, **Nizar Habash** and Christof Monz. Use of Minimal Lexical Conceptual Structures for Single-Document Summarization. University of Maryland technical report #LAMP-TR-113, CAR-TR-997, CS-TR-4596, UMIACS-TR-2004-39, 2004.
- [4] **Habash, Nizar**. Generation-Heavy Hybrid Machine Translation. Doctoral Dissertation. Computer Science Department, University of Maryland College Park, 2003.
- [5] **Nizar Habash**. A Reference Manual to the Linearization Engine oxyGen. University of Maryland technical report: LAMP-TR-079, CS-TR-4295, UMIACS-TR-2001-73, MDA-904-96-C-1250. 2001.
- [6] Habash, Nizar and Jin Tong. MFTV: Multi-Faceted Tree Viewer. University of Maryland Computer Science Department Technical Report: CS-TR-4528. 1999.

Other Publications

- [1] Interviewed for an article on machine translation in **Wired Magazine (Issue 14.12)** - December 2006)
- [2] **Habash, Nizar**. "On Arabic and its Dialects," Multilingual Magazine. #81 Volume 17 Issue 5, 2006.
- [3] **Habash, Nizar**. Introduction to Delason: The Complete Guide to the Artificial Language. Self-published book manuscript, 1998. URL: <http://www.delason.com>

Oral Presentations

Invited Talks

- [1] **Habash, Nizar**. Arabic Computational Linguistics: Past, Present and Future. Invited Talk. Arabic Linguistics Symposium. Columbia University, March 2012.
- [2] **Habash, Nizar**. Arabic Morphological Analysis and Disambiguation: Problem, Solution, Applications. Invited Talk. Linguistics Speaker Series. Georgetown University, Washington D.C., 2009.
- [3] **Habash, Nizar**. Automatic Diacritization of Arabic Text. Invited Talk. Columbia Linguistic Society, Columbia University, New York, 2009.
- [4] **Habash, Nizar**. Arabic Diacritization through Full Morphological Tagging. Invited Talk, Language Technology Institute Seminar Series. Carnegie Mellon University, 2007.
- [5] **Habash, Nizar**. Arabic Dialect Modeling: from Morphological Analysis to Parsing. Invited Talk, NSF Funded US-Morocco Workshop on Language Technology Research and Education. Ecole Nationale de l'Industrie Minérale, Rabat, Morocco, 2006.
- [6] **Habash, Nizar**. Disambiguation of Rich Arabic Morphological Analyses. Invited Talk, NYU Natural Language Processing Colloquium. April 14, 2006.
- [7] Mona Diab, **Nizar Habash**, and Owen Rambow. Arabic Dialect Parsing. Invited Talk, Linguistic Data Consortium, University of Pennsylvania, December 9, 2005
- [8] **Habash, Nizar**. Combining Symbolic and Statistical Techniques for Machine Translation and Morphological Disambiguation. Invited Talk, AT&T Labs, Florham Park, NJ, October 7, 2005.
- [9] **Habash, Nizar**. Generation-Heavy Hybrid Machine Translation. Invited Talk. IBM TJ Watson Research Center, March 9, 2005.
- [10] **Habash, Nizar**. Generation-Heavy Hybrid Machine Translation. Invited Talk. Computer Science Department Seminar, City College of New York, March 8, 2005.
- [11] **Habash, Nizar**. Generation-Heavy Hybrid Machine Translation. Invited Talk. Language Technologies Institute Seminar. Carnegie Mellon University, November 19, 2004.
- [12] **Habash, Nizar**. Semitic Linguistic Phenomena. Invited Talk. Workshop on Machine Translation for Semitic Languages, MT Summit, New Orleans, LA, 2003.

Invited Panels

- [1] **Habash, Nizar**. Treebanking for Machine Translation. Invited Panel Talk, GALE PI Meeting, Tampa, Florida, 2008.
- [2] **Habash, Nizar**. NLG is still relevant to MT. Invited Panel Talk. The MT Summit XI Workshop on Using corpora for natural language generation: language generation and machine translation (UCNLG+MT), Copenhagen, Denmark, September, 2007.
- [3] **Habash, Nizar**. Hybrid Machine Translation. Invited Panel Talk. Conference of the Association for Machine Translation in the Americas, AMTA, 2006.

Invited Tutorials

- [1] **Habash, Nizar**. Introduction to Arabic Natural Language Processing. A four-hour tutorial, 25 attendees, Carnegie Mellon Qatar, Doha, Qatar, 2010.

- [2] Diab, Mona and **Nizar Habash**. Introduction to Arabic Computational Linguistics: Modern Standard Arabic and Arabic Dialects. A four-hour tutorial. 40 attendees. In GURT: Georgetown University Round Table conference on Arabic Language and Linguistics, Washington D.C., 2010.
- [3] **Habash, Nizar**. Introduction to Arabic Natural Language Processing. A three-hour tutorial, 35 attendees. Johns Hopkins University Summer School on Human Language Technology, Baltimore, 2006.
- [4] **Habash, Nizar**. Introduction to Arabic Natural Language Processing. A three-hour tutorial, 40 attendees. BBN Technologies, Boston, Massachusetts, 2005.
- [5] **Habash, Nizar**. Introduction to Arabic Natural Language Processing. A three-hour tutorial, 20 attendees. The National Cryptologic Museum, 2005.
- [6] **Habash, Nizar**. Introduction to Arabic Natural Language Processing. A three-hour tutorial, 50 attendees. Johns Hopkins University Summer Workshop, Baltimore, Maryland, 2005.

Colloquia, Forums, Workshops, Bake-offs

- [1] Marine Carpuat, Mahmoud Ghoneim, **Habash, Nizar** and Mona Diab. Columbia University 2009 Arabic-English MT Evaluation Systems. NIST Machine Translation Evaluation, 2009.
- [2] Diab, Mona, **Nizar Habash** and Owen Rambow. Tools for Arabic Morphological Analysis, Disambiguation and Tokenization. Foreign Language Technology Forum, July 2008.
- [3] **Habash, Nizar**. Techniques for Online OOV Handling in Phrase-based Statistical Machine Translation. *Columbia University NLP Colloquium*, April 2008.
- [4] **Habash, Nizar**. Five Techniques for OOV Reduction in Arabic-English and Urdu-English SMT. *NIST MTEval Workshop*, March 2008.
- [5] **Habash, Nizar**, Mahmoud Ghoneim, Hayden Metsky and Mona Diab. Columbia University 2008 Arabic-English and Urdu-English MT Evaluation Systems. NIST Machine Translation 2008 (MT-08) Evaluation, March, 2008.
- [6] **Habash, Nizar**, Halim Abbas, Bonnie Dorr, Christof Monz and Necip Ayan. Columbia University 2006 Arabic-English MT Evaluation Systems. NIST Machine Translation 2006 (MT-06) Evaluation. September, 2006.
- [7] David Chiang, Bonnie Dorr, **Nizar Habash**, Christof Monz and Philip Resnik, and. The University of Maryland College Park 2005 Chinese-English and Arabic-English MT Evaluation Systems. NIST Machine Translation 2005 (MT-05) Evaluation, June, 2005.
- [8] **Habash, Nizar**. Workshop Task Description and Results. AMTA'04 Seventh Interlingua Workshop Determining Interlingua Utility for Machine Translation Georgetown University, Washington DC, October 2, 2004.
- [9] Kumar, Shankar, Yonggang Deng, Charles Schafer, Woosung Kim, Paola Virga, **Nizar Habash**, David Smith, Filip Jurcicek, Bill Byrne, Sanjeev Khudanpur, Zak Shafran, and David Yarowsky. The Johns Hopkins University 2004 Chinese-English and Arabic-English MT Evaluation Systems. NIST Machine Translation 2004 (MT-04) Evaluation, June 22, 2004.

Available Software and Resources

Tools

- [1] MADA+TOKAN: Morphological Analysis and Disambiguation for Arabic (2005, developed with Owen Rambow and Ryan Roth, current version 3.1)
This suite includes other tools that I developed:
TOKAN: General Arabic Tokenizer.
ALMOR: Arabic Lexeme-based Morphological Generator/Analyzer.
[<http://www1.ccls.columbia.edu/MADA/>]
- [2] MAGEAD: Morphological Analyzer/Generator for Arabic and its Dialects (2006, developed with Owen Rambow, current version 0.5)
Includes MorphTools, an extension to the Lextools/ATT FSM toolkit to handle multi-tier morphology
[http://www1.ccls.columbia.edu/~habash/software-downloads/MAGEAD_Distribution.html]

- [3] REMOOV: Tool for Reducing Out-Of-Vocabulary in MT (2007)
[<http://www1.ccls.columbia.edu/~habash/REMOOV-0.4.tar.gz>]
- [4] Arabic Constituency-to-Dependency Conversion tool (2009)
[<http://www1.ccls.columbia.edu/~habash/C2D-v0.6-4Mar2011.tar.gz>]

Resources

- [5] Annotated Corpus of Arabic Functional Number, Gender and Rationality (2011) [forthcoming]
- [6] CATiB v1.0: Columbia Arabic Treebank (2009) [<http://www1.ccls.columbia.edu/~CATiB/>]
- [7] CatVar A Categorical Variation Database for English (2002)
[<http://clipdemos.umiacs.umd.edu/catvar/>]

Service

Doctoral Committees

- [1] Weiyun Ma, Columbia University, 2012 (candidacy)
- [2] Fadi Biadisy, Columbia University, 2008 (candidacy), 2010 (proposal), 2011 (defense).
- [3] Kristen Parton, Columbia University, 2009 (candidacy), 2010 (proposal).
- [4] Jakob Elming, Copenhagen Business School, 2009 (defense).
- [5] Josep Crego, Universitat Politècnica de Catalunya, 2008 (defense).
- [6] Otakar Smrz, Charles University, Czech Republic, 2007 (defense).
- [7] Andrew Rosenberg, Columbia University, 2006 (candidacy).

Masters Reader

- [1] Ibrahim Saleh, Georgetown University, Washington, DC, 2009.

Research Proposal Reviews

- [1] Review panel member, American Association for the Advancement of Science (AAAS) (2010, 2011).
- [2] Qatar National Research Fund reviewer (2009, 2010).
- [3] NSF review panel (2009).
- [4] NSF proposal reviewer (2005, 2006).

Journal Editor/Editorial Boards

- [1] Member of the editorial board of Computational Linguistics Journal (2012-2014)
- [2] Co-editor with Hany Hassan for a special issue of MT journal on Arabic Machine Translation, 2012.
- [3] Member of the editorial board of the Journal for Machine Translation (2007-)

Journal Reviewing

- [1] Journal of Computational Linguistics, (2003, 2006, 2010, 2011).
- [2] Journal of Natural Language Engineering, (2004, 2007, 2011).
- [3] Journal of Pattern Recognition Research (2011).
- [4] Journal of Artificial Intelligence Research (2011).
- [5] ACM Transaction on Asian Languages Information Processing (TALIP), (2004, 2010).
- [6] IEEE Transactions on Audio, Speech and Language Processing (2004, 2005, 2006, 2007).

Conference/Workshop Program Committees

- [1] Conference of the Association for Computational Linguistics (ACL): 2008-2006, 2002
- [2] Conference of the Association for Machine Translation in the Americas (AMTA): 2012-2006
- [3] Conference of the European ACL (EACL): 2012, 2006
- [4] Conference of the North American ACL (NAACL): 2007, 2006, 2001
- [5] Conference on Computational Linguistics (COLING): 2008, 2004,
- [6] Conference on Empirical Methods for Natural Language Processing (EMNLP): 2009-2007
- [7] Conference on Mediterranean Arabic Language and Speech Technology (MEDAR): 2009
- [8] International Conference Arabic Language Technology (ALTIC): 2011
- [9] International Conference on the Arabic Language Processing (CITALA): 2007
- [10] International Natural Language Generation Conference (INLG): 2002
- [11] Language Resources and Evaluation Conference (LREC): 2012, 2008, 2006
- [12] Machine Translation Summit (MT Summit): 2011, 2007
- [13] Theoretical and Methodological Issues in Machine Translation (TMI): 2007
- [14] Workshop on Computational Approaches to Arabic Script Languages: 2009, 2004
- [15] Workshop on Computational Approaches to Semitic Languages: 2007, 2005
- [16] Workshop on Statistical Machine Translation (WMT): 2011, 2010

Professional Organizations

- [1] Secretary of the board of the International Association for Machine Translation (IAMT) (2009-present)
- [2] Secretary of the board of the Association for Machine Translation in the Americas (AMTA) (2008-present)
- [3] Research community representative in the board of the Association for Machine Translation in the Americas (AMTA) (2006-2008)
- [4] Vice president of the ACL Special Interest Group on Semitic Languages (2006-2009)

Conference and Workshop Organization

- [1] Publication co-chair for NAACL 2012.
- [2] Workshop program co-chair for NAACL 2009.
- [3] Senior Program Committee member for NAACL 2007 (machine translation track).
- [4] Program co-chair for the conference of the Association for Machine Translation in the Americas (AMTA'06)
- [5] One of three organizers for a workshop on Computational Approaches to Semitic Languages ACL 2005.
- [6] Chair of the Workshop on Determining Interlingua Utility for Machine Translation. AMTA-2004, Washington DC, 2004.
- [7] One of three organizers of the Workshop on Machine Translation for Semitic Languages, MT Summit, New Orleans, LA, 2003.

Advisory Committees

- [1] Member of the Qatar Computational Research Center (QCRI) Arabic Language Technology Roundtable. Doha, Qatar, January, 2011.
- [2] Co-chair of the advisory committee to improving the quality of the Arabic Treebank, 2006-2008.
- [3] Member of an advisory committee to NIST MT Eval program, 2007.
- [4] Member of the machine translation team under the NSF funded MINDS workshop, 2006.

Honors

- [1] Best Paper Award from the Traitement Automatique du Langage Naturel (TALN) conference, Montréal, Canada, 2010.
- [2] Graduate Student Service Award, University of Maryland College Park, 2003.

- [3] The Samuel N. Alexander Fellowship, Association for Computing Machinery, Washington D.C. Chapter, 2001.
- [4] The Phi Kappa Phi Award of Excellence, Phi Kappa Phi, 1997/1998.
- [5] Selected to represent Old Dominion University as one of five graduating students featured in a newspaper article "[Pomp & Circumstance](#)" The Virginian-Pilot, May 9, 1997.
- [6] The College of Engineering and Technology Outstanding Scholar Award, Old Dominion University, 1997.
- [7] The Faculty Award in Computer Engineering, College of Engineering and Technology, Old Dominion University, 1997.
- [8] The Outstanding Individualized Study Student Award, College of Arts and Letters, Old Dominion University, 1997.
- [9] The Award of Academic Excellence, Academic Honors Program Old Dominion University, 1997.
- [10] The Meghan O'Connor Award for Academic Achievement and Community Service, Academic Honors Program, Old Dominion University, 1997.
- [11] Undergraduate Research Award, Academic Honors Program, Old Dominion University, 1994.

References

Available upon request