

Nizar Habash

NYU Abu Dhabi, Saadiyat Island
PO Box 129188, Abu Dhabi, UAE
Office: +971-2-628-4578
<http://www.nizarhabash.com>
nizar@nizarhabash.com

Education

Ph.D. Computer Science, University of Maryland College Park, 2003.

Dissertation title: Generation-Heavy Hybrid Machine Translation

Advisor: Prof. Bonnie Dorr

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

Sep. 2014 to present. New York University Abu Dhabi, Computer Science Department, Associate Professor.

Sep. 2014 to present. Columbia University, Center for Computational Learning Systems, Adjunct Research Scientist.

Jul. 2014 to present. Birzeit University, Sina Institute for Knowledge Engineering and Arabic Technologies, Adjunct Associate Professor.

Feb. 2009 to Aug. 2014. Columbia University, Center for Computational Learning Systems, Research Scientist.

Feb. 2009 to Aug. 2014. 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. Cypcorp, Inc., Intern Ontological Engineer.

Research Experience

Research Areas

- My primary research area is natural language processing (NLP) / computational linguistics (CL).
- My specialty sub-areas are Arabic NLP, computational morphology and machine translation.
- I co-founded the Columbia Arabic Dialect Modeling (CADIM) group in 2005 with Owen Rambow and Mona Diab.
- I founded the Computational Approaches to Modeling Language (CAMEL) Lab in NYU Abu Dhabi in 2015.

Citation Metrics

- Citations 5,549; h-index 39; i10-index 117. Google Scholar (June 30, 2017).

Past and Current Research Support

- [1] "Morphological Annotation of Gulf Arabic." Nizar Habash (PI). New York University Abu Dhabi Research Enhancement Fund. Awarded \$40,000. 2016.
- [2] "SAMER: Simplification of Arabic Masterpieces for Extensive Reading." Muhamed Al Khalil (PI), Nizar Habash (co-PI) and Dris Soulaïmani (co-PI). New York University Abu Dhabi Research Enhancement Fund. Awarded \$181,000. Two-year project. 2016.
- [3] "Workshop on Arabic Dialect Technologies." Nizar Habash (PI). New York University Abu Dhabi Institute Funding. 2016.
- [4] "MADAR: Multi-Arabic Dialect Applications and Resources." Nizar Habash (PI), Kemal Oflazer, Owen Rambow and Houda Bouamor. Qatar National Research Fund Award. Approved. \$866,551 total spread over three years. Starts Feb 2016.
- [5] "A Multilingual-to-Arabic Machine Translation Evaluation Dataset." Nizar Habash (PI). New York University Abu Dhabi Research Enhancement Fund. Awarded \$40,000. 2015.
- [6] "Opinion Mining for Arabic with Models of Semantics and Credibility." Hazem Hajj (PI), Khaled Shaban, Wassim El Hajj, Osama Halabi and Nizar Habash. Qatar National Research Fund Award. Approved. \$331,000 per year for three years starting in November 2013.
- [7] "Arabic Dialect Morphological Analysis." Nizar Habash (PI), Mona Diab and Owen Rambow. LDC Subcontract. DARPA BOLT Program. Year 1 (\$88,142), Year 2 (\$91,443) and Year 3 (\$90,000) (10/2011-6/2015) approved.
- [8] "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, 2 and 3 (10/2011-3/2015) approved.
- [9] "Automatic Correction of Standard Arabic Text: Resource and System Development." Nizar Habash (PI), Kemal Oflazer and Behrang Mohit. Qatar National Research Fund Award. Approved. \$327,333 per year for three years starting in 2012.
- [10] "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.
- [11] "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 6/2012-2/2013 (\$153,661).
- [12] "Natural Language Processing Tools for Dialectal Arabic (NADIA)." Mona Diab (PI), Nizar Habash and Owen Rambow. AXCIOM subcontract. Technical Support Working Group. 4/2012-12/2013 (\$900,000).
- [13] "Preserving Morphological Richness in Poor-Pivot-based Statistical Machine Translation." Nizar Habash (PI). Google Research Award. 9/2012-8/2013 (\$84,211).
- [14] "Improving Statistical Machine Translation into Arabic." Nizar Habash (PI). Science Applications International Corporation (SAIC) research grant. 1/2012-8/2013 (\$120,307).
- [15] "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).
- [16] "Arabic Dialect Tools and Resources for COLABA." Mona Diab (PI), Nizar Habash and Owen Rambow. Axciom Subcontract. TSWG COLABA Program. P1 10/2008-3/2010 (\$442,709); P2 4/2010-3/2011(\$257,618)
- [17] "Arabic Dialect Translation for Agile." Nizar Habash (PI). IBM Subcontract. DARPA GALE Program. GALE-Y5 6/2010-6/2011 (\$112,437).
- [18] "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).
- [19] "Columbia Arabic Tree Bank." Nizar Habash (PI). IBM Subcontract. DARPA GALE Program. GALE-Y3 4/2008-8/2008 (\$200,000).
- [20] "SGER: Automatic Processing of Natural Language Code Switching." Mona Diab (PI), Nizar Habash and Owen Rambow. NSF. 9/2007-2/2009 (\$40,467).
- [21] "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).

- [22] "CINDOR: Conceptual Interlingua Document Retrieval." Owen Rambow (PI) and Nizar Habash (Post-doc). Textwise, Inc. 2/2005-5/2006 (\$38,534).
- [23] "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).
- [24] "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 have been the primary advisor of the following doctoral students.

- [1] Ahmed El Kholy, Computer Science Department, Columbia University. 2008-2015. Currently with Microsoft Research.
- [2] Wael Salloum, Computer Science Department, Columbia University. 2010-present.
- [3] Nasser Zalmout, Computer Science, New York University. 2015-present
- [4] Dima Taji, Computer Science, New York University. 2017-present

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

Senior Colleagues

- [5] Dr. Jim Jing-Yan Wang, Research Scientist. Computational Approaches to Modeling Language Lab, New York University, Abu Dhabi, 2016-present.
- [6] Dr. Hieu Hoang, Research Scientist. Computational Approaches to Modeling Language Lab, New York University, Abu Dhabi, 2015.
- [7] Dr. Wajdi Zaghouni, Lead Annotator, Carnegie Mellon University Qatar (with Kemal Oflazer, Behrang Mohit and Houda Bouamor on two projects). 2011-present.

Post-doctoral

- [8] Alla Rozovskaya, Post-doctoral researcher, Columbia University Center for Computational Learning Systems. 2013-present.
- [9] Thomas Lippincott, Post-doctoral researcher, Columbia University Center for Computational Learning Systems (with Owen Rambow). 2013-present.
- [10] Nadi Tomeh, Post-doctoral researcher, Columbia University Center for Computational Learning Systems. (Primary research advisor). 2012-2013.
- [11] Abdelati Hawwari, Post-doctoral researcher, Columbia University Center for Computational Learning Systems (with Mona Diab and Owen Rambow). 2011-2012.
- [12] Marine Carpuat, Post-doctoral researcher, Columbia University Center for Computational Learning Systems (with Mona Diab and Owen Rambow). 2010-2011.
- [13] Yuval Marton, Post-doctoral researcher, Columbia University Center for Computational Learning Systems. (Primary research advisor). 2009-2011.
- [14] Ryan Roth, Post-doctoral researcher, Columbia University Center for Computational Learning Systems (with Mona Diab and Owen Rambow). 2008-2009.

Doctoral

- [15] Sarah Alkuhlani, Computer Science Department, Columbia University. 2009-2015.
- [16] Mohammad Rasooli, Ph.D. student, Computer Science Department, Columbia University. (Advisor: Nizar Habash 2012-2014; Michael Collins 2014-graduation).
- [17] Noura Farra, Ph.D. student, Computer Science Department, Columbia University. (Advisor: Nizar Habash 2012-2014; Kathy McKeown 2014-graduation).
- [18] Kristen Parton, Ph.D. student, Computer Science Department, Columbia University. (Advisor: Kathy McKeown). 2010-2012.
- [19] Kais Dukes, Ph.D. student, University of Leeds. (Advisor: Eric Atwell; Secondary: Nizar Habash). 2009-present.
- [20] Mohamed El-Tantawy, Ph.D. student, Computer Science Department, Columbia University. (Advisor: Owen Rambow). 2008-2011.

- [21] Fadi Biadsy, Ph.D. student, Computer Science Department, Columbia University. (Advisor: Julia Hirschberg). 2008-2011.
- [22] Josep M Crego. Ph.D. student, Computer Science Department, Universitat Politècnica de Catalunya. (Advisor: José Mariño). 2007-2009
- [23] Jakob Elming, Ph.D. student, Computer Science Department, Copenhagen Business School. (Advisor: Daniel Hardt). 2006-2009.
- [24] Mahmoud Ghoneim. Ph.D. student, Computer Science Department, Cairo University. (Columbia Advisor: Mona Diab). 2006-2009.
- [25] Necip Fazil Ayan, Ph.D. student. Computer Science Department, University of Maryland College Park. (Advisor: Bonnie Dorr). 2003-2005.

Junior Colleagues

- [26] Hind Saddiki, Research Assistant. Co-advised with Prof. Muhammad Al-Khalil. New York University, Abu Dhabi, 2016-now.
- [27] Fadhl Eryani, Research Assistant. Computational Approaches to Modeling Language Lab, New York University, Abu Dhabi, 2017-now.
- [28] Alexander Erdmann, Visiting Doctoral Student. Computational Approaches to Modeling Language Lab, New York University, Abu Dhabi, 2016-present.
- [29] Sara Hassan, Research Assistant. Computational Approaches to Modeling Language Lab, New York University, Abu Dhabi, 2016-2017.
- [30] Dima Taji, Research Assistant. Computational Approaches to Modeling Language Lab, New York University, Abu Dhabi, 2015-2017.
- [31] Anas Shahrour, Research Assistant. Computational Approaches to Modeling Language Lab, New York University, Abu Dhabi, 2015-2016.
- [32] Salam Khalifa, Research Assistant. Computational Approaches to Modeling Language Lab, New York University, Abu Dhabi, 2015-present.
- [33] Ossama Obeid, Research Programmer, Carnegie Mellon University Qatar (with Behrang Mohit). 2011-present.
- [34] Ryan Roth, Staff Associate, Columbia University Center for Computational Learning Systems (with Mona Diab and Owen Rambow). 2010-2013.
- [35] Ramy Eskander, Staff Associate, Columbia University Center for Computational Learning Systems (with Owen Rambow). 2012-present.
- [36] Arfath Pasha, Staff Associate, Columbia University Center for Computational Learning Systems (with Owen Rambow and Mona Diab). 2012-present.
- [37] Mohamed Badrashiny, Staff Associate, Columbia University Center for Computational Learning Systems (with Owen Rambow and Mona Diab). 2012-2013.
- [38] Sakhar ElKhereyf. Computer Research Institute, King Abdulaziz City for Science and Technology (KACST). Fall 2013. Independent Researcher.

Masters

- [39] Sara Alwaalan. M.S. student, Computer Science Department, Columbia University. Fall 2013.
- [40] Amit Abbi. M.S. student, Computer Science Department, Columbia University (with Owen Rambow and Mona Diab). 2010-2011.
- [41] Reshef Shilon, M.S. student, Tel Aviv University. (Advisor: Shuly Wintner). 2010-2011.
- [42] Jun Hu, M.S. student, Computer Science Department, Columbia University. (Primary research advisor). 2008-2009.
- [43] Ryan Roth, M.S. student, Computer Science Department, Columbia University (with Owen Rambow). 2007.
- [44] Abdelhalim Abbas, M.S. student, Computer Science Department, Columbia University. (Primary research advisor). 2006.
- [45] Farhad Peydaye Saheli, M.S. student, Computer Science Department, Columbia University (with Owen Rambow). 2006.
- [46] Fadi Biadsy, M.S. student, Computer Science Department, Ben-Gurion University. (Advisor: Jihad El-Sana). 2005-2006.
- [47] Matthew Chu, M.S. student, Computer Science Department, Columbia University (with Owen Rambow). 2005.

Undergraduate

- [48] Daniel Watson, undergraduate student, Computer Science, NYUAD, Summer 2017-present.
- [49] Talha Javed, undergraduate student, Computer Science, NYUAD, Spring-Summer 2017. (Capstone Advisor)
- [50] Dana Abu Ali, undergraduate student, Computer Science, NYUAD, Summer 2016-present. (Capstone Advisor)
- [51] William Held, undergraduate student, Computer Science, NYUAD, Spring 2016-present.
- [52] Hayat Abual Hassan, undergraduate student, Computer Science, NYUAD, Spring 2017- now. (Capstone Advisor)
- [53] Jihyun Kim, undergraduate student, Computer Science, NYUAD, Spring 2017- now. (Capstone Advisor)
- [54] Jose Varias, undergraduate student, Computer Science, NYUAD, Spring 2017- now. (Capstone Advisor)
- [55] Ming Hu, undergraduate student, Computer Science, NYUAD, Spring 2017- now. (Capstone Advisor)
- [56] Muaz Ahmad, undergraduate student, Computer Science, NYUAD, Spring 2017- now. (Capstone Advisor)
- [57] Pauka Dozsa, undergraduate student, Computer Science, NYUAD, Spring 2017- now. (Capstone Advisor)
- [58] Shehroze Khan, undergraduate student, Computer Science, NYUAD, Spring 2017- now. (Capstone Advisor)
- [59] Tarik Zulfikarpasic, undergraduate student, Computer Science, NYUAD, Spring 2017- now. (Capstone Advisor)
- [60] Wei Chen, undergraduate student, Computer Science, NYUAD, Spring 2017- now. (Capstone Advisor)
- [61] Morgante Pell, undergraduate student, Computer Science, NYUAD, Fall 2015-Spring 2016. (Capstone Advisor)
- [62] Asyrique "Ash" Thevendran, undergraduate student, Computer Science, NYUAD, Fall 2015-now. (Capstone Advisor)
- [63] Chukwuyem "Chuksy" Onyibe, undergraduate student, Computer Science, NYUAD, Fall 2015-now. (Capstone Advisor)
- [64] Lingliang "Ling" Zhang, undergraduate student, Computer Science, NYUAD, Fall 2015-Fall 2016. (Capstone Advisor)
- [65] Jermsak "Mick" Jermsurawong, undergraduate student, Computer Science, NYUAD, Fall 2014-Spring 2015. (Capstone Advisor)
- [66] Arvind Srinivasan, undergraduate student, Computer Science Department, Columbia University. Fall 2013.
- [67] Matt Capetola, undergraduate student, Computer Science Department, Columbia University (with Owen Rambow). 2010-2011.
- [68] Jacob Andreas, undergraduate student, Computer Science Department, Columbia University (with Owen Rambow). 2010-2011.
- [69] Morgan Rhodes, undergraduate student, Computer Science Department, Columbia University. 2006-2007.
- [70] Amitai Aviram, undergraduate student, Computer Science Department, Columbia University (with Owen Rambow). 2005.

High School

- [71] 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.

Publications

Journal Articles

- [1] Baly, Ramy, Hazem Hajj, **Nizar Habash**, Khaled Bashir Shaban, and Wassim El-Hajj. A Sentiment Treebank and Morphologically Enriched Recursive Deep Models for Effective Sentiment Analysis in Arabic. In Transactions of Asian and Low-Resource Language Information Processing (TALLIP). Accepted. 2017.
- [1] Zalmout, Nasser and **Nizar Habash**. Optimizing Tokenization Choice for Machine Translation across Multiple Target Languages. The Prague Bulletin of Mathematical Linguistics No. 108, (2017), pp. 257–269. doi: 10.1515/pralin-2017-0025. [*Duplicate listing in this CV as Journal and Conference Proceedings*]
- [2] Jarrar, Mustafa, **Nizar Habash**, Faeq Alrimawi, Diyam Akra and Nasser Zalmout. Curras: An Annotated Corpus for the Palestinian Arabic Dialect. *Language Resources & Evaluation* (2016). doi:10.1007/s10579-016-9370-7.
- [3] Salloum, Wael and **Nizar Habash**. ADAM: Analyzer for Dialectal Arabic Morphology. Journal of King Saud University - Computer and Information Sciences. Volume 26, Issue 4, December 2014, Pages 372–378.
- [4] Arts, Tressy, Yonatan Belinkov, **Nizar Habash**, Adam Kilgarriff, and Vit Suchomel. arTenTen: Arabic Corpus and Word Sketches. Journal of King Saud University - Computer and Information Sciences. Volume 26, Issue 4, December 2014, Pages 357–371.
- [5] Marton, Yuval, **Nizar Habash** and Owen Rambow. “Dependency Parsing of Modern Standard Arabic with Lexical and Inflectional Features”. *Computational Linguistics, Volume: 39, Issue: 1, 2013*.
- [6] Kais Dukes, Eric Atwell and **Nizar Habash**. “Supervised Collaboration for Syntactic Annotation of Quranic Arabic”, *Language Resource and Evaluation Journal*, 47:33-62, 2013.
- [7] 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*, 26:25-45, 2012.
- [8] Marine Carpuat, Yuval Marton and **Nizar Habash**. “Improved Arabic-to-English statistical machine translation by reordering post-verbal subjects for word alignment”. In *Machine Translation Journal Special Issue on Arabic Machine Translation*, 26:105-120, 2012.
- [9] Shilon, Reshef, **Nizar Habash**, Alon Lavie, Shuly Wintner. “Machine Translation between Hebrew and Arabic”. In *Machine Translation Journal Special Issue on Arabic Machine Translation*, 26:177-195, 2012.
- [10] 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, 16(3): 197-243, 2010.
- [11] **Habash, Nizar**, Bonnie J. Dorr and Chirstof Monz. Symbolic-to-Statistical Hybridization: Extending Generation-Heavy Machine Translation. *Machine Translation Journal*, 23:23--63, 2009.
- [12] **Habash, Nizar**, Bonnie Dorr, and David Traum. Hybrid Natural Language Generation from Lexical Conceptual Structures. *MT Journal* volume 18 (2): 81-128, 2003.
- [13] 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(2):118-123, 2003.
- [14] **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.
This book was translated to Arabic by Prof. Hend Al-Khalifa of King Saud University in Riyadh, Saudi Arabia (2014).

Book Chapters

- [1] Fabri, Ray, Michael Gasser, **Nizar Habash**, George Kiraz, and Shuly Wintner. "Chapter 1: Linguistic Introduction: The Orthography, Morphology and Syntax of Semitic Languages." In Zitouni, Imed, ed., "Semitic Natural Language Processing," Springer, 2014. Pages 3-41.
- [2] **Nizar Habash** and Fatiha Sadat. "Arabic Preprocessing for Statistical Machine Translation: Schemes, Techniques and Combinations." In Souidi, A., Neumann, G., Farghaly, A., Zbib, R. eds, "Challenges for Arabic MT", Chapter 5, John Benjamins NLP series, 2012.
- [3] 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](#) (Section 1.4.2). In the Handbook of Natural Language Processing and Machine Translation, Joseph Olive, Caitlin Christianson and John McCary, eds. Springer, 2011.
- [4] Maamouri, Mohamed, Ann Bies, Seth Kulick, **Habash, Nizar**, Reem Faraj and Ryan Roth. "Arabic Treebanking." In [Chapter 1 on Data Acquisition and Linguistic Resources](#) (Section 1.3.4). In the Handbook of Natural Language Processing and Machine Translation, Joseph Olive, Caitlin Christianson and John McCary, eds. Springer, 2011.
- [5] 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](#) (Section 2.2.1). In the Handbook of Natural Language Processing and Machine Translation, Joseph Olive, Caitlin Christianson and John McCary, eds. Springer, 2011.
- [6] **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.
- [7] **Habash, Nizar**. "Using Surface Span in Syntax-aware MT Evaluation (SEPIA)." In [Chapter 5 on Machine Translation Evaluation and Optimization](#) (Section 5.6.4). In the Handbook of Natural Language Processing and Machine Translation, Joseph Olive, Caitlin Christianson and John McCary, eds. Springer, 2011.
- [8] 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. "Interlingual Annotation of Multilingual Text Corpora and FrameNet." In: H.C. Boas (ed.), /Multilingual FrameNets in Computational Lexicography: Methods and Applications/, Chapter 10, 287-318. Berlin: Mouton de Gruyter, 2009.
- [9] 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.
- [10] **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 Souidi. 2007.
- [11] **Habash, Nizar**, Abdelhadi Souidi, and Tim Buckwalter. "On Arabic Transliteration." Book Chapter. In Arabic Computational Morphology: Knowledge-based and Empirical Methods. Editors Antal van den Bosch and Abdelhadi Souidi. 2007.

Conference Proceedings

2017

- [1] **Habash, Nizar**, Nasser Zalmout, Dima Taji, Hieu Hoang, and Maverick Alzate. A Parallel Corpus for Evaluating Machine Translation between Arabic and European Languages. In Proceedings of the European Chapter of the Association for Computational Linguistics (EACL), Valencia, Spain, 2017.
- [2] Zalmout, Nasser and **Nizar Habash**. Optimizing Tokenization Choice for Machine Translation across Multiple Target Languages. In Proceedings of the European Association of Machine Translation. Prague, Czech Republic, 2017.
- [3] Zeman, Daniel, Martin Popel, Milan Straka, Jan Hajic, Joakim Nivre, Filip Ginter, Juhani Luotolahti, Sampo Pyysalo, Slav Petrov, Martin Potthast, Francis Tyers, Elena Badmaeva, Memduh Gokirmak, Anna Nedoluzhko, Silvie Cinkova, Jan Hajic jr., Jaroslava Hlavacova, Václava Kettnerová, Zdenka Uresova, Jenna Kanerva, Stina Ojala, Anna Missilä, Christopher D. Manning, Sebastian Schuster, Siva Reddy, Dima Taji, **Nizar Habash**, Herman Leung, Marie-Catherine de Marneffe, Manuela Sanguinetti, Maria Simi, Hiroshi Kanayama, Valeria dePaiva, Kira Drohanova, Héctor Martínez Alonso, Çağrı Çöltekin, Umut Sulubacak, Hans Uszkoreit, Vivien Macketanz, Aljoscha Burchardt, Kim Harris, Katrin Marheinecke, Georg Rehm, Tolga Kayadelen, Mohammed Attia, Ali Elkahky, Zhuoran Yu, Emily Pitler, Saran Lertpradit, Michael Mandl, Jesse Kirchner, Hector Fernandez Alcalde, Jana Strnadová, Esha Banerjee, Ruli Manurung, Antonio Stella, Atsuko Shimada, Sookyoung Kwak, Gustavo Mendonca, Tatiana Lando, Rattima Nitisaroj and Josie Li. CoNLL 2017 Shared Task: Multilingual Parsing from Raw Text to Universal Dependencies. In Proceedings of the Conference on Computational Natural Language Learning (CoNLL). Vancouver, Canada, 2017.

2016

- [4] Abu Ali, Dana and **Nizar Habash**. Botta: An Arabic Dialect Chatbot. In Proceedings of the International Conference on Computational Linguistics (COLING): System Demonstrations, Osaka, Japan, 2016.
- [5] Eskander, Ramy, **Nizar Habash**, Owen Rambow and Arfath Pasha. Creating Resources for Dialectal Arabic from a Single Annotation: A Case Study on Egyptian and Levantine. In Proceedings of the International Conference on Computational Linguistics (COLING), Osaka, Japan, 2016.
- [6] Guzmán, Francisco, Houda Bouamor, Ramy Baly and **Nizar Habash**. Machine Translation Evaluation for Arabic using Morphologically-Enriched Embeddings. In Proceedings of the International Conference on Computational Linguistics (COLING), Osaka, Japan, 2016.
- [7] Salam Khalifa, Nasser Zalmout and **Nizar Habash**. YAMAMA: Yet Another Multi-Dialect Arabic Morphological Analyzer. In Proceedings of the International Conference on Computational Linguistics (COLING): System Demonstrations, Osaka, Japan, 2016.
- [8] Shahrour, Anas, Salam Khalifa, Dima Taji and **Nizar Habash**. CamelParser: A System for Arabic Syntactic Analysis and Morphological Disambiguation. In Proceedings of the International Conference on Computational Linguistics (COLING): System Demonstrations, Osaka, Japan, 2016.
- [9] Al Zaatari, Ayman, Reem El Ballouli, Shady Elbassouni, Wassim El-Hajj, Hazem Hajj, Khaled Shaban and **Nizar Habash**. Arabic Corpora for Credibility Analysis. In Proceedings of the Language Resources and Evaluation Conference (LREC), Portorož, Slovenia, 2016.
- [10] Al-Badrashiny, Mohamed, Arfath Pasha, Mona Diab, **Nizar Habash**, Owen Rambow, Wael Salloum and Ramy Eskander. SPLIT: Smart Preprocessing (Quasi) Language Independent Tool. In Proceedings of the Language Resources and Evaluation Conference (LREC), Portorož, Slovenia, 2016.
- [11] Al-Shargi, Faisal, Aidan Kaplan, Ramy Eskander, **Nizar Habash** and Owen Rambow. A Morphologically Annotated Corpus and a Morphological Analyzer for Moroccan and San'ani

Yemeni Arabic. In Proceedings of the Language Resources and Evaluation Conference (LREC), Portorož, Slovenia, 2016.

- [12] **Habash, Nizar**, Anas Shahrour and Muhamed Al-Khalil. Exploiting Arabic Diacritization for High Quality Automatic Annotation. In Proceedings of the Language Resources and Evaluation Conference (LREC), Portorož, Slovenia, 2016.
- [13] Khalifa, Salam, Houda Bouamor and **Nizar Habash**. DALILA: The Dialectal Arabic Linguistic Learning Assistant. In Proceedings of the Language Resources and Evaluation Conference (LREC), Portorož, Slovenia, 2016.
- [14] Khalifa, Salam, **Nizar Habash**, Dana Abdulrahim and Sara Hassan. A Large Scale Corpus of Gulf Arabic. In Proceedings of the Language Resources and Evaluation Conference (LREC), Portorož, Slovenia, 2016.
- [15] Temnikova, Irina, Wajdi Zaghouni, Stephan Vogel and **Nizar Habash**. Applying the Cognitive Machine Translation Evaluation Approach to Arabic. In Proceedings of the Language Resources and Evaluation Conference (LREC), Portorož, Slovenia, 2016.
- [16] Zaghouni, Wajdi, **Nizar Habash**, Ossama Obeid, Behrang Mohit and Kemal Oflazer. Building an Arabic Machine Translation Post-Edited Corpus: Guidelines and Annotation. In Proceedings of the Language Resources and Evaluation Conference (LREC), Portorož, Slovenia, 2016.

2015

- [17] El Kholly, Ahmed, and **Nizar Habash**. Morphological Constraints for Phrase Pivot Statistical Machine Translation. In Proceedings of Machine Translation Summit. Miami, 2015.
- [18] Shahrour, Anas, Salam Khalifa and **Nizar Habash**. Improving Arabic Diacritization through Syntactic Analysis. In Proceedings of EMNLP, Lisbon, 2015.
- [19] Jermsurawong, Jermsak and **Nizar Habash**. Predicting the structure of Cooking Recipes. In Proceedings of EMNLP, Lisbon, 2015.
- [20] Masmoudi, Abir, **Nizar Habash**, Mariem Ellouze, Yannick Estève, Lamia Hadrich Belguith. Arabic Transliteration of Romanized Tunisian Dialect Text: A Preliminary Investigation. In Proceedings of CICLING-2015, Cairo, 2015.

2014

- [21] Sikos, Jennifer, Peter David, Nizar Habash, and Reem Faraj. Authorship Analysis of Inspire Magazine through Stylometric and Psychological Features. Joint Intelligence and Security Informatics Conference (JISIC) Proceedings, September 2014.
- [22] Cui, Xiaodong, Brian Kingsbury, Jia Cui, Bhuvana Ramabhadran, Andrew Rosenberg, Mohammed Sadegh Rasooli, Owen Rambow, **Nizar Habash** and Vaibhava Goel. Improving Deep Neural Network Acoustic Modeling For Audio Corpus Indexing Under The IARPA Babel Program. In Proceedings of INTERSPEECH 2014, Singapore, 2014.
- [23] El Kholly, Ahmed and **Nizar Habash**. Alignment Symmetrization Optimization targeting Phrase Pivot Statistical Machine Translation. In Proceedings of the Conference of the European Association for Machine Translation (EAMT2014), Dubrovnik, Croatia, 2014.
- [24] Al-Badrashiny, Mohamed, Ramy Eskander, **Nizar Habash** and Owen Rambow. Automatic Transliteration of Romanized Dialectal Arabic. In Proceedings of the Conference on Computational Natural Language Learning (CONLL), Baltimore, Maryland, 2014.
- [25] Rozovskaya, Alla, Kai-Wei Chang, Mark Sammons, Dan Roth and **Nizar Habash**. The Illinois-Columbia System in the CoNLL-2014 Shared Task. In Proceedings of the Conference on Computational Natural Language Learning Shared Task (CONLL-ST), Baltimore, Maryland, 2014.
- [26] Salloum, Wael, Heba Elfardy, Linda Alamir Salloum, **Nizar Habash** and Mona Diab. Sentence Level Dialect Identification for Machine Translation System Selection. In Proceedings of the Conference of the Association for Computational Linguistics (ACL), Baltimore, Maryland, 2014.

- [27] Farra, Noura, Nadi Tomeh, Alla Rozovskaya, and **Nizar Habash**. Generalized Character-Level Spelling Error Correction. In Proceedings of the Conference of the Association for Computational Linguistics (ACL), Baltimore, Maryland, 2014.
- [28] Rasooli, Mohammad Sadegh, **Nizar Habash**, Owen Rambow and Thomas Lippincot. Unsupervised Morphology-Based Vocabulary Expansion. In Proceedings of the Conference of the Association for Computational Linguistics (ACL), Baltimore, Maryland, 2014.
- [29] Bouamor, Houda, **Nizar Habash** and Kemal Oflazer. A Multidialectal Parallel Corpus of Arabic. In Proceedings of the Language Resources and Evaluation Conference (LREC), Reykjavik, Iceland, 2014.
- [30] Diab, Mona, Mohamed Al-Badrashiny, Maryam Aminian, Mohammed Attia, Pradeep Dasigi, Heba Elfardy, Ramy Eskander, **Nizar Habash**, Abdelati Hawwari, Wael Salloum. Tharwa: A Large Scale Dialectal Arabic - Standard Arabic - English Lexicon. In Proceedings of the Language Resources and Evaluation Conference (LREC), Reykjavik, Iceland, 2014.
- [31] Maamouri, Mohamed, Ann Bies, Seth Kulick, Michael Ciul, **Nizar Habash** and Ramy Eskander. Developing an Egyptian Arabic Treebank: Impact of Dialectal Morphology on Annotation and Tool Development. In Proceedings of the Language Resources and Evaluation Conference (LREC), Reykjavik, Iceland, 2014.
- [32] Masmoudi, Abir, Mariem Ellouze Khmekhem, Yannick Esteve, Lamia Hadrach Belguith and **Nizar Habash**. A Corpus and Phonetic Dictionary for Tunisian Arabic Speech Recognition. In Proceedings of the Language Resources and Evaluation Conference (LREC), Reykjavik, Iceland, 2014.
- [33] Pasha, Arfath, Mohamed Al-Badrashiny, Mona Diab, Ahmed El Kholy, Ramy Eskander, **Nizar Habash**, Manoj Pooleery, Owen Rambow and Ryan M. Roth. MADAMIRA: A Fast, Comprehensive Tool for Morphological Analysis and Disambiguation of Arabic. In Proceedings of the Language Resources and Evaluation Conference (LREC), Reykjavik, Iceland, 2014.
- [34] Zaghouani, Wajdi, Behrang Mohit, **Nizar Habash**, Ossama Obeid, Nadi Tomeh, Alla Rozovskaya, Noura Farra, Sarah Alkuhlani and Kemal Oflazer. Large Scale Arabic Error Annotation: Guidelines and Framework. In Proceedings of the Language Resources and Evaluation Conference (LREC), Reykjavik, Iceland, 2014.
- [35] Zribi, Ines, Rahma Boujelbane, Abir Masmoudi, Mariem Ellouze, Lamia Belguith and **Nizar Habash**. A Conventional Orthography for Tunisian Arabic. In Proceedings of the Language Resources and Evaluation Conference (LREC), Reykjavik, Iceland, 2014.

2013

- [36] Eskander, Ramy, **Nizar Habash** and Owen Rambow. Automatic Extraction of Morphological Lexicons from Morphologically Annotated Corpora. In Proceedings of the conference on Empirical Methods for Natural Language Processing (EMNLP), Seattle, WA, 2013.
- [37] Hamdi, Ahmed, Rahma Boujelbane, **Nizar Habash** and Alexis Nasr. The Effects of Factorizing Root and Pattern Mapping in Translating between Tunisian Arabic and Standard Arabic. In Proceedings of the Machine Translation Summit (MTSummit), Nice, France, 2013.
- [38] El Kholy, Ahmed, **Nizar Habash**, Gregor Leusch, Evgeny Matusov, Hassan Sawaf. Selective Combination of Pivot and Direct Statistical Machine Translation Models. In Proceedings of the International Joint Conference on Natural Language Processing (IJCNLP), Nagoya, Japan, 2013.
- [39] Rasooli, Mohammad Sadegh, Ahmed El Kholy and **Nizar Habash**. Orthographic and Morphological Processing for Persian-to-English Statistical Machine Translation. In Proceedings of the International Joint Conference on Natural Language Processing (IJCNLP), Nagoya, Japan, 2013.
- [40] Obeid, Ossama, Wajdi Zaghouani, Behrang Mohit, **Nizar Habash**, Kemal Oflazer and Nadi Tomeh. A Web-based Annotation Framework For Large-Scale Text Correction. Demo Paper. In

Proceedings of the International Joint Conference on Natural Language Processing (IJCNLP), Nagoya, Japan, 2013.

- [41] Pasha, Arfath, Mohammad Al-Badrashiny, Mona T. Diab, **Nizar Habash**, Manoj Pooleery, Owen Rambow, Ryan M. Roth and Mohamed Altantawy. DIRA: Dialectal Arabic Information Retrieval Assistant. Demo Paper. In Proceedings of the International Joint Conference on Natural Language Processing (IJCNLP), Nagoya, Japan, 2013.
- [42] El Kholy, Ahmed, **Nizar Habash**, Gregor Leusch, Evgeny Matusov, Hassan Sawaf. Language Independent Connectivity Strength Features for Phrase Pivot Statistical Machine Translation. In Proceedings of Conference of the Association for Computational Linguistics (ACL), Sofia, Bulgaria, 2013.
- [43] Tomeh, Nadi, **Nizar Habash**, Ryan Roth, Noura Farra, Pradeep Dasigi and Mona Diab. Reranking with Linguistic and Semantic Features for Arabic Optical Character Recognition. In Proceedings of Conference of the Association for Computational Linguistics (ACL), Sofia, Bulgaria, 2013.
- [44] **Habash, Nizar**, Ryan Roth, Owen Rambow, Ramy Eskander and Nadi Tomeh. Morphological Analysis and Disambiguation for Dialectal Arabic. In Proceedings of Conference of the North American Association for Computational Linguistics (NAACL), Atlanta, Georgia, 2013.
- [45] Alkuhlani, Sarah, **Nizar Habash** and Ryan Roth. Automatic Morphological Enrichment of a Morphologically Underspecified Treebank. In Proceedings of Conference of the North American Association for Computational Linguistics (NAACL), Atlanta, Georgia, 2013.
- [46] Salloum, Wael and **Nizar Habash**. Dialectal Arabic to English Machine Translation: Pivoting through Modern Standard Arabic. In Proceedings of Conference of the North American Association for Computational Linguistics (NAACL), Atlanta, Georgia, 2013.
- [47] Eskander, Ramy, **Nizar Habash**, Owen Rambow and Nadi Tomeh. Processing Spontaneous Orthography. In Proceedings of Conference of the North American Association for Computational Linguistics (NAACL), Atlanta, Georgia, 2013.
- [48] Hamdi, Ahmed, Rahma Boujelbane, **Nizar Habash** and Alexis Nasr. Un système de traduction de verbes entre arabe standard et arabe dialectal par analyse morphologique profonde. In Proceedings of Traitement Automatique du Langage Naturel 2013 (TALN 2013), Les Sables-d'Olonne, France, 2013.

2012

- [49] Salloum, Wael and **Nizar Habash**. Elissa: A Dialectal to Standard Arabic Machine Translation System. Demo Paper. In Proceedings of the International Conference on Computational Linguistics (COLING). Mumbai, India, 2012.
- [50] Parton, Kristen, **Nizar Habash**, and Kathleen McKeown. Lost and Found in Translation: The Impact of Machine Translated Results on Translingual Information Retrieval. In Proceedings of the conference of the Association for Machine Translation in the Americas (AMTA), San Diego, 2012.
- [51] El Kholy, Ahmed, **Nizar Habash**. Rich Morphology Generation Using Statistical Machine Translation. In Proceedings of the International Natural Language Generation Conference (INLG), Starved Rock, IL, USA, 2012.
- [52] El Kholy, Ahmed, **Nizar Habash**. Translate, Predict or Generate: Modeling Rich Morphology in Statistical Machine Translation. In Proceedings of the Conference of the European Association for Machine Translation (EAMT), Trento, Italy, 2012.
- [53] Singh, Nimesh, **Nizar Habash**. Hebrew Morphological Preprocessing for Statistical Machine Translation. In Proceedings of the Conference of the European Association for Machine Translation (EAMT), Trento, Italy, 2012.
- [54] Parton, Kristen, **Nizar Habash**, Kathleen McKeown, Gonzalo Iglesias, Adrià de Gispert. Can Automatic Post-Editing Make MT More Meaningful? In Proceedings of the Conference of the European Association for Machine Translation (EAMT), Trento, Italy, 2012.
- [55] Alkuhlani, Sarah and **Nizar Habash**. Identifying Broken Plurals, Irregular Gender, and Rationality in Arabic Text. In Proceedings of Conference of the European Association for Computational Linguistics (EACL), Avignon, France, 2012.

- [56] **Habash, Nizar**, Mona Diab and Owen Rambow. Conventional Orthography for Dialectal Arabic. In Proceedings of International Conference on Language Resources and Evaluation (LREC), Istanbul, 2012.

2011

- [57] 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.
- [58] 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.
- [59] 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.
- [60] **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.
- [61] 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.

2010

- [62] 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.
- [63] 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.
- [64] 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.
- [65] Altantawy, Mohamed, **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.
- [66] Kais Dukes and **Nizar Habash**. Morphological Annotation of Quranic Arabic. In Proceedings of the Language Resources and Evaluation Conference (LREC), Malta, 2010.

2009

- [67] **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.
- [68] Biadsy, 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.
- [69] **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.
- [70] **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.
- [71] **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.

2008

- [72] **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.
- [73] **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.
- [74] 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.
- [75] **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.
- [76] 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.

2007

- [77] **Habash, Nizar**. Syntactic Preprocessing for Statistical Machine Translation. In Proceedings of the Machine Translation Summit (MT-Summit), Copenhagen, Denmark, 2007.
- [78] 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.
- [79] 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.
- [80] **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.
- [81] **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.
- [82] 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.
- [83] **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.

2006

- [84] **Habash, Nizar**, Bonnie Dorr and Christof Monz. Challenges in Building an Arabic-English GHMT System with SMT Components. In Proceedings of the Association for Machine Translation in the Americas (AMTA-2006), Boston, MA, 2006.
- [85] **Habash, Nizar** and Owen Rambow. MAGEAD: A Morphological Analyzer and Generator for the Arabic Dialects. In Proceedings of ACL, Sydney, Australia, 2006.
- [86] Sadat, Fatiha and **Nizar Habash**. Combination of Arabic Preprocessing Schemes for Statistical Machine Translation. In Proceedings of ACL, Sydney, Australia, 2006.
- [87] **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.
- [88] Chiang, David, Mona Diab, **Nizar Habash**, Owen Rambow, and Safi Shareef. Parsing Arabic Dialect. In Proceedings of the European chapter of the Association of Computational Linguistics (EACL). 2006.

- [89] **Habash, Nizar**, Clinton Mah, Randy Calistri-Yeh, Sabiha Imran and Paraic Sheridan. Design, Construction and Validation of an Arabic-English Conceptual Interlingua for Cross-lingual Information Retrieval. In Proceedings of the International Conference on Language Resources and Evaluation (LREC), 2006.
- [90] 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.
- [91] 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.
- [92] 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.

2005

- [93] **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.

2004

- [94] **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.
- [95] **Habash, Nizar**. Large Scale Lexeme Based Arabic Morphological Generation. In Proceedings of Traitement Automatique du Langage Naturel (TALN-04). Fez, Morocco, 2004.
- [96] **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.
- [97] Ayan, Fazil, Bonnie J. Dorr, and **Nizar Habash**. Multi-Align: Combining Linguistic and Statistical Techniques to Improve Alignments for Adaptable MT. In Proceedings of the 6th Conference of the Association for Machine Translation in the Americas (AMTA-2004), Washington DC, 2004.
- [98] 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), Washington DC, 2004.

2003

- [99] **Habash, Nizar**. Matador: A Large Scale Spanish-English GHMT System. In Proceedings of the MT Summit, New Orleans, LA, 2003.
- [100] **Habash, Nizar**. System Demonstration Matador: Spanish-English GHMT. In Proceedings of the MT Summit, New Orleans, LA, 2003.
- [101] **Habash, Nizar** and Bonnie Dorr. A Categorical Variation Database for English. In Proceedings of North American Association for Computational Linguistics, Edmonton, Canada, 2003.
- [102] **Habash, Nizar** and Bonnie Dorr. System Demonstration CatVar: A Database of Categorical Variation for English. In Proceedings of the MT Summit, New Orleans, LA, 2003.

2002

- [103] **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.

- [104] 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.
- [105] **Habash, Nizar**. Generation-Heavy Hybrid Machine Translation. In Proceedings of the International Natural Language Generation Conference (NLG). New York, 2002.

2001-1998

- [106] **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.
- [107] **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.
- [108] **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.
- [109] **Habash, Nizar**. Issues in Palestinian Arabic Spelling Standardization. In the North American Conference on Afro-Asiatic Linguistics (NACAL 27), Baltimore, MD, 1999.
- [110] 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

2017

- [1] Onyibe, Chukwuyem and Nizar Habash. OMAM at SemEval-2017 Task 4: English Sentiment Analysis with Conditional Random Fields. International Workshop on Semantic Evaluation. ACL. Vancouver, 2017.
- [2] Baly, Ramy, Gilbert Badaro, Ali Hamdi, Rawan Moukalled, Rita Aoun, Georges El-Khoury, Ahmad Al Sallab, Hazem Hajj, Nizar Habash, Khaled Shaban and Wassim El-Hajj. OMAM at SemEval-2017 Task 4: Evaluation of English State-of-the-Art Sentiment Analysis Models for Arabic and a New Topic-based Model. International Workshop on Semantic Evaluation. ACL. Vancouver, 2017.
- [3] Taji, Dima, Nizar Habash, and Daniel Zeman. Universal Dependencies for Arabic. In Proceedings of the Workshop on Arabic Natural Language Processing (WANLP 2017), (co-located with EACL 2017), Valencia, Spain, 2017.
- [4] Khalifa, Salam, Sara Hassan and Nizar Habash. A Morphological Analyzer for Gulf Arabic Verbs. In Proceedings of the Workshop on Arabic Natural Language Processing (WANLP 2017), (co-located with EACL 2017), Valencia, Spain, 2017.
- [5] Baly, Ramy, Gilbert Badaro, Georges El-Khoury, Rawan Moukalled, Rita Aoun, Hazem Hajj, Wassim El-Hajj, Nizar Habash and Khaled Shaban. A Characterization Study of Arabic Twitter Data with a Benchmarking for State-of-the-Art Opinion Mining Models. In Proceedings of the Workshop on Arabic Natural Language Processing (WANLP 2017), (co-located with EACL 2017), Valencia, Spain, 2017.
- [6] Zhang, Lingliang, Nizar Habash and Godfried Toussaint. Robust Dictionary Lookup in Multiple Noisy Orthographies. In Proceedings of the Workshop on Arabic Natural Language Processing (WANLP 2017), (co-located with EACL 2017), Valencia, Spain, 2017.

2016

- [7] Taji, Dima, Ramy Eskander, Nizar Habash, and Owen Rambow. The Columbia University-New York University Abu Dhabi SIGMORPHON 2016 Morphological Reinflection shared task submission. SIGMORPHON Workshop, ACL, Berlin, 2016.
- [8] Zaghouni, Wajdi, Nizar Habash, Houda Bouamor, Ossama Obeid, Sawsan Alqahtani, Mona Diab and Kemal Oflazer. Filtering Dialectal Arabic Text in Two Large Scale Annotation Projects. In COLING 2016 Workshop on Noisy User-generated Text (W-NUT), Osaka, Japan, 2016.
- [9] Zalmout, Nasser, Hind Saddiki and Nizar Habash. Analysis of Foreign Language Teaching Methods: An Automatic Readability Approach. In Proceedings of the Workshop on Natural Language Processing Techniques for Educational Applications, Osaka, Japan, 2016.

2015

- [10] Saadane, Houda, and **Nizar Habash**. A Conventional Orthography for Algerian Arabic. In Proceedings of the Second Arabic Natural Language Processing Workshop, ACL, Beijing, 2015.
- [11] Hamdi, Ahmed, Alexis Nasr, **Nizar Habash**, and Núria Gala. POS-tagging of Tunisian Dialect Using Standard Arabic Resources and Tools. In Proceedings of the Second Arabic Natural Language Processing Workshop, ACL, Beijing, 2015.
- [12] Rozovskaya, Alla, Houda Bouamor, **Nizar Habash**, Wajdi Zaghouni, Ossama Obeid, Behrang Mohit. The Second QALB Shared Task on Automatic Text Correction for Arabic. In Proceedings of the Second Arabic Natural Language Processing Workshop, ACL, Beijing, 2015.
- [13] Farra, Nourra, Kathleen McKeown, **Nizar Habash**. Annotating Targets of Opinions in Arabic using Crowdsourcing. In Proceedings of the Second Arabic Natural Language Processing Workshop, ACL, Beijing, 2015.
- [14] Zaghouni, Wajdi, **Nizar Habash**, Houda Bouamor, Alla Rozovskaya, Behrang Mohit, Abeer Heider, and Kemal Oflazer. Correction Annotation for Non-Native Arabic Texts: Guidelines and Corpus. In Proceedings of The 9th Linguistic Annotation Workshop, NAACL, pp. 129-139. 2015.
- [15] Sajjad, Hassan, Nadir Durrani, Francisco Guzman, Preslav Nakov, Ahmed Abdelali, Stephan Vogel, Wael Salloum, Ahmed El Kholy, **Nizar Habash**. QCN System Description for NIST OpenMT15. In Proceedings of the Open Machine Translation Evaluation Workshop, Washington DC, 2015.

2014

- [16] Eskander, Ramy, Mohamed Al-Badrashiny, **Nizar Habash** and Owen Rambow. Foreign Words and the Automatic Processing of Arabic Social Media Text Written in Roman Script. In Proceedings of the First Workshop on Computational Approaches to Code Switching, EMNLP, Doha, 2014.
- [17] Badaro, Gilbert, Ramy Baly, Hazem Hajj, **Nizar Habash** and Wassim El-Hajj. A Large Scale Arabic Sentiment Lexicon for Arabic Opinion Mining. In Proceedings of the Arabic Natural Language Processing Workshop, EMNLP, Doha, 2014.
- [18] Bies, Ann, Zhiyi Song, Mohamed Maamouri, Stephen Grimes, Haejoong Lee, Jonathan Wright, Stephanie Strassel, **Nizar Habash**, Ramy Eskander, and Owen Rambow. Transliteration of Arabizi into Arabic Orthography: Developing a Parallel Annotated Arabizi-Arabic Script SMS/Chat Corpus. In Proceedings of the Arabic Natural Language Processing Workshop, EMNLP, Doha, 2014.
- [19] Jarrar, Mustafa, Nizar Habash, Diyam Akra and Nasser Zalmout. Building a Corpus for Palestinian Arabic: a Preliminary Study. In Proceedings of the Arabic Natural Language Processing Workshop, EMNLP, Doha, 2014.
- [20] Jeblee, Serena, Weston Feely, Houda Bouamor, Alon Lavie, **Nizar Habash** and Kemal Oflazer. Domain and Dialect Adaptation for Machine Translation into Egyptian Arabic. In Proceedings of the Arabic Natural Language Processing Workshop, EMNLP, Doha, 2014.

- [21] Mohit , Behrang, Alla Rozovskaya, **Nizar Habash**, Wajdi Zaghouni and Ossama Obeid. The First shared Task on Automatic Text Correction for Arabic. In Proceedings of the Arabic Natural Language Processing Workshop, EMNLP, Doha, 2014.
- [22] Rozovskaya, Alla, **Nizar Habash**, Ramy Eskander, Noura Farra and Wael Salloum. The Columbia System in the QALB-2014 Shared Task on Arabic Error Correction. In Proceedings of the Arabic Natural Language Processing Workshop, EMNLP, Doha, 2014.
- [23] Tomeh, Nadi, **Nizar Habash**, Ramy Eskander and Joseph Le Roux. A Pipeline Approach to Supervised Automatic Error Correction. In Proceedings of the Arabic Natural Language Processing Workshop, EMNLP, Doha, 2014.

2013

- [24] Djamé Seddah, Reut Tsarfaty, Sandra Kübler, Marie Candito, Jinho D. Choi, Richárd Farkas, Jennifer Foster, Iakes Goenaga, Koldo Gojenola, Yoav Goldberg, Spence Green, **Nizar Habash**, Marco Kuhlmann, Wolfgang Maier, Joakim Nivre, Adam Przepiorkowski, Ryan Roth, Wolfgang Seeker, Yan- nick Versley, Veronika Vincze, Marcin Woliński, Alina Wróblewska, and Eric Villemonte de la Clérgerie. 2013. Overview of the SPMRL 2013 Shared Task: A Cross-Framework Evaluation of Parsing Morphologically Rich Languages. In Proceedings of the 4th Workshop on Statistical Parsing of Morphologically Rich Languages: Shared Task, SPMRL'13, EMNLP, Seattle, 2013.
- [25] Marton, Yuval, **Nizar Habash**, Owen Rambow and Sarah Alkuhlani. SPMRL'13 Shared Task System: The CADIM Arabic Dependency Parser. In Proceedings of the 4th Workshop on Statistical Parsing of Morphologically Rich Languages: Shared Task, SPMRL'13, EMNLP, Seattle, 2013.
- [26] Eskander, Ramy, **Nizar Habash**, Ann Bies, Seth Kulick, and Mohamed Maamouri. Automatic Correction and Extension of Morphological Annotations. In the Proceedings of the Linguistic Annotations Workshop (LAW) in the Conference of the Association for Computational Linguistics (ACL), Sofia, Bulgaria, 2013.
- [27] Belinkov, Yonatan, **Nizar Habash**, Adam Kilgarriff, Noam Ordan, Ryan Roth and Vitek Suchomel. arTenTen: a new, vast corpus for Arabic. In Proceedings of WACL'2 Workshop on Arabic Corpus Linguistics. Lancaster, UK, 2013.

2012

- [28] **Habash, Nizar**, Ramy Eskander and Abdelati Hawwari. A Morphological Analyzer for Egyptian Arabic. In the Proceedings of the Workshop on Computational Research in Phonetics, Phonology, and Morphology (SIGMORPHON) in the Conference of the North American chapter of the Association for Computational Linguistics (NAACL), Montreal, Canada, 2012.

2011

- [29] Andreas, Jacob, **Nizar Habash**, and Owen Rambow. Fuzzy syntactic reordering for phrase-based statistical machine translation. In Proceedings of the Workshop on Statistical Machine Translation at the Conference for Empirical Methods in Natural Language Processing (EMNLP 2011), Edinburgh, Scotland, UK, 2011.
- [30] Marton, Yuval, 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 at the Conference for Empirical Methods in Natural Language Processing (EMNLP 2011), Edinburgh, Scotland, UK, 2011.
- [31] 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.

- [32] 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 (FSMNL 2011), Blois, France, 2011.

2010

- [33] 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.
- [34] Kholy, 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.
- [35] 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.
- [36] 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.

2009

- [37] 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.
- [38] **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.
- [39] Biadsy, 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.
- [40] 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.
- [41] 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.

2008

- [42] **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.
- [43] 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.

- [44] **Habash, Nizar**, Owen Rambow, Mona Diab and Reem Faraj. Guidelines for Annotation of Arabic Dialectness. In Proceedings of workshop on Arabic and its local languages (LREC), Marrakech, Morocco, 2008.

2006-2000

- [45] Biadisy, Fadi, Jihad El-Sana and **Nizar Habash**. Online Arabic Handwriting Recognition Using Hidden Markov Models. International Workshop on Handwriting and Optical Character Recognition, Paris, France, 2006.
- [46] **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.
- [47] Farwell, David, Stephen Helmreich, Bonnie J. Dorr, **Nizar Habash**, Florence Reeder, Keith Miller, Lori Levin, Teruko Mitamura, Eduard Hovy, Owen Rambow, and Advaith 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.
- [48] 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 Advaith Siddharthan. Semantic Annotation for Interlingual Representation of Multilingual Text. In Proceedings of the Workshop on Beyond Named Entity Recognition: Semantic Labeling for NLP Tasks, LREC, Portugal, 2004.
- [49] 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 Advaith Siddharthan. Semantic Annotation and Lexico-Syntactic Paraphrase. In Proceedings of the Workshop on Building Lexical Resources from Semantically Annotated Corpora, LREC, Portugal, 2004.
- [50] 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.
- [51] Traum, David and **Nizar Habash**. Generation from Lexical Conceptual Structures. Workshop on Applied Interlinguas, ANLP/NAACL-2000. Seattle, WA.

Technical Reports

- [1] Almahairi, Amjad, Kyunghyun Cho, **Nizar Habash**, and Aaron Courville. First result on Arabic neural machine translation. arXiv preprint arXiv:1606.02680 (2016).
- [2] **Habash, Nizar** and Sarah Alkuhlani. Annotation Guidelines for Arabic Nominal Gender, Number, and Rationality. Center for Computational Learning Systems technical report #CCLS-13-03. 2013.
- [3] Diab, Mona, **Nizar Habash**, Owen Rambow and Ryan M. Roth. LDC Arabic Treebanks and Associated Corpora: Data Divisions Manual. Center for Computational Learning Systems technical report #CCLS-13-02. 2013.
- [4] Salloum, Wael and **Nizar Habash**. Elissa: A Dialectal to Standard Arabic Machine Translation System. Center for Computational Learning Systems technical report #CCLS-12-04. 2012.
- [5] Salloum, Wael and **Nizar Habash**. A Modern Standard Arabic Closed-Class Word List. Center for Computational Learning Systems technical report #CCLS-12-03. 2012.
- [6] Mohamed Maamouri, Sondos Krouna, Dalila Tabessi, Nadia Hamrouni, and **Nizar Habash**. Egyptian Arabic Morphological Annotation Guidelines. Linguistic Data Consortium Technical Report. 2012
- [7] **Habash, Nizar**, Mona Diab, and Owen Rambow. CODA: Conventional Orthography for Dialectal Arabic. Center for Computational Learning Systems technical report #CCLS-11-02. 2011.
- [8] **Habash, Nizar**, Owen Rambow and Ryan Roth. MADA+TOKAN Manual. Center for Computational Learning Systems technical report #CCLS-10-01. 2010.

- [9] 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.
- [10] **Habash, Nizar**. Generation-Heavy Hybrid Machine Translation. Doctoral Dissertation. Computer Science Department, University of Maryland College Park, 2003.
- [11] **Habash, Nizar**. 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.
- [12] **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 Wamda.com. “Exploring the frontiers of Arabic natural language processing”. May 2016. <http://www.wamda.com/memakersge/2016/05/exploring-frontiers-arabic-natural-language-processing>
- [2] Interview with The Gazelle, the NYUAD Magazine. “The art of artificial languages”. March 6, 2016. <http://www.thegazelle.org/issue/issue-81/features/nizar-habash/>
- [3] **Habash, Nizar**. An Introduction to Semitish. Self-published book manuscript, 2015. <http://www.nizarhabash.com/palisra/SemitishPrimer-rtfplus.pdf>
- [4] Interviewed for an article on machine translation for Arabic speakers (in Arabic) on **Radio Sawa Website (April 19, 2013)**
<http://www.radiosawa.com/content/how-to-use-google-translate-arabic-bing/222331.html>
- [5] Interviewed for an article on machine translation in **Wired Magazine (Issue 14.12 - December 2006)**
- [6] **Habash, Nizar**. “On Arabic and its Dialects,” Multilingual Magazine. #81 Volume 17 Issue 5, 2006.
- [7] **Habash, Nizar**. Introduction to Delason: The Complete Guide to the Artificial Language. Self-published book manuscript, First edition 1998. Second edition 2012.
URL: http://www.nizarhabash.com/delason/delason_digest.pdf

Oral Presentations

Invited Talks

- [1] **Habash, Nizar**. Computational Processing of Arabic for Digital Humanistic Research. Digital Humanities Abu Dhabi DHAD conference. 12 April 2017.
- [2] **Habash, Nizar**. Language Technologies for Arabic and its Dialects. NYUAD Institute Public talk. November 15, 2016.
- [3] **Habash, Nizar**. Computational Processing of the Arabic Language. Invited Talk. Nazarbayev University, Astana, Kazakhstan, October 20, 2016.
- [4] **Habash, Nizar**. Computational Processing of Arabic Dialects: Challenges, Advances & Future Directions. Keynote Talk. The 2nd Workshop on Arabic Corpora and Processing Tools, LREC-2016, Portorož, Slovenia, May 24, 2016.
- [5] **Habash, Nizar**. Morphological Processing of Arabic and its Dialects. Invited Talk. University of Malta, Malta, April 22, 2016.
- [6] **Habash, Nizar**. Computational Processing of Gulf Arabic. Invited Talk. British University of Dubai, Dubai, UAE, April 13, 2016.
- [7] **Habash, Nizar**. Computational Processing of Arabic Dialects. Invited Talk. International Workshop on Computers and Information Sciences (WCIS 2015). Tabuk, Saudi Arabia, October 8, 2015. (Talk cancelled due to technical difficulties, but slides were part of the proceedings).
- [8] **Habash, Nizar**. Computational Processing of Arabic Dialects. Invited Talk. Université Toulouse III - Paul Sabatier., Toulouse, France, May 11, 2015.

- [9] **Habash, Nizar.** Arabic and Arabic Dialect Technologies. Invited Speaker. NYU Abu Dhabi Tech Fest, April 25, 2015.
- [10] **Habash, Nizar.** Arabic Language Technology: Current Status and Future Potential. Invited Speaker Defining Standards Conference, Abu Dhabi, UAE, November 19, 2014.
- [11] **Habash, Nizar.** Morphological Processing of Arabic and its Dialects. Computer Science Seminar Series. NYU Abu Dhabi, November 13, 2014.
- [12] **Habash, Nizar.** Computational Processing of Arabic Dialects. Keynote Talk. Workshop on Language Technology for Closely Related Languages and Language Variants (LT4CloseLang), EMNLP, Doha, October 29, 2014.
- [13] **Habash, Nizar.** Morphological Analysis and Disambiguation for Dialectal Arabic. Invited talk. IBM Professional Interest Community Seminar, June 26, 2013.
- [14] **Habash, Nizar.** Arabic Natural Language Processing: Challenges, Solutions, Applications. Invited talk. New York University Abu Dhabi, May 2, 2013.
- [15] **Habash, Nizar.** Arabic Computational Linguistics: Challenges and Opportunities. Keynote Presentation. The 2nd Palestinian Symposium on Computational Linguistics and Arabic Content (iArabic'2013). Sina Institute for Knowledge Engineering and Arabic Technologies. Bir Zeit University. April 20, 2013.
- [16] **Habash, Nizar.** Arabic Natural Language Processing: Challenges, Solutions, Applications. Invited talk. Invited talk. Polytechnic Institute of New York University, April 11, 2013.
- [17] **Habash, Nizar.** Arabic Natural Language Processing: Challenges, Solutions, Applications. Invited talk. Invited talk. Courant Institute, New York University, April 10, 2013.
- [18] **Habash, Nizar.** Arabic Natural Language Processing: Challenges, Solutions, Applications. Invited talk. Invited talk. Language Technologies Institute. Carnegie Mellon University, April 4, 2013.
- [19] **Habash, Nizar.** Arabic Natural Language Processing: Challenges, Solutions, Applications. Invited talk. Invited talk. Computer and Information Technology Department, Purdue University, April 2, 2013.
- [20] **Habash, Nizar.** Arabic Dialect Processing and Speech-to-Speech Machine Translation. Workshop on Simultaneous Interpretation and Multimodal Translation. Invited talk. ATT, March 25, 2013.
- [21] **Habash, Nizar.** Orthographic and Morphological Processing for English-Arabic Statistical Machine Translation. Keynote Presentation, Arabic Language Technology Forum. Carnegie Mellon University Qatar and Qatar Computational Research Institute, Doha, Qatar, November 2012.
- [22] **Habash, Nizar.** Arabic Parsing and Statistical Machine Translation. Invited Talk, Language Technology Institute. Carnegie Mellon University, July 2012.
- [23] **Habash, Nizar.** Arabic Computational Linguistics: Past, Present and Future. Keynote Presentation. Arabic Linguistics Symposium. Columbia University, March 2012.
- [24] **Habash, Nizar.** Arabic Morphological Analysis and Disambiguation: Problem, Solution, Applications. Invited Talk. Linguistics Speaker Series. Georgetown University, Washington D.C., 2009.
- [25] **Habash, Nizar.** Automatic Diacritization of Arabic Text. Invited Talk. Columbia Linguistic Society, Columbia University, New York, 2009.
- [26] **Habash, Nizar.** Arabic Diacritization through Full Morphological Tagging. Invited Talk, Language Technology Institute Seminar Series. Carnegie Mellon University, 2007.
- [27] **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.
- [28] **Habash, Nizar.** Disambiguation of Rich Arabic Morphological Analyses. Invited Talk, NYU Natural Language Processing Colloquium. April 14, 2006.
- [29] **Mona Diab, Nizar Habash,** and Owen Rambow. Arabic Dialect Parsing. Invited Talk, Linguistic Data Consortium, University of Pennsylvania, December 9, 2005
- [30] **Habash, Nizar.** Combining Symbolic and Statistical Techniques for Machine Translation and Morphological Disambiguation. Invited Talk, AT&T Labs, Florham Park, NJ, October 7, 2005.
- [31] **Habash, Nizar.** Generation-Heavy Hybrid Machine Translation. Invited Talk. IBM TJ Watson Research Center, March 9, 2005.

- [32] **Habash, Nizar.** Generation-Heavy Hybrid Machine Translation. Invited Talk. Computer Science Department Seminar, City College of New York, March 8, 2005.
- [33] **Habash, Nizar.** Generation-Heavy Hybrid Machine Translation. Invited Talk. Language Technologies Institute Seminar. Carnegie Mellon University, November 19, 2004.
- [34] **Habash, Nizar.** Semitic Linguistic Phenomena. Invited Talk. Workshop on Machine Translation for Semitic Languages, MT Summit, New Orleans, LA, 2003.

Invited Posters

- [1] Taji, Dima, **Nizar Habash**, and Daniel Zeman. Universal Dependencies for Arabic. Invited Poster at NoDaLiDa Workshop on Universal Dependencies (UDW 2017), Gothenburg, Sweden, 2017.

Invited Panels

- [1] **Habash, Nizar.** Future Challenges for Arabic Resource Development. Invited Panelist. Workshop on Free/Open-Source Arabic Corpora and Corpora Processing Tools, in the Language Resources and Evaluation Conference (LREC), Reykjavik, 2014
- [2] **Habash, Nizar.** Treebanking for Machine Translation. Invited Panel Talk, GALE PI Meeting, Tampa, Florida, 2008.
- [3] **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.
- [4] **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] Zaghouani, Wajdi, Nizar Habash, Ossama Obeid, Behrang Mohit, Houda Bouamor, Kemal Oflazer. Annotation Guidelines and Framework for Arabic Machine Translation Post-Edited Corpus. In Qatar Foundation Annual Research Conference Proceedings Doha, Qatar, 2016.
- [2] Ahmed El Kholy and Nizar Habash. English Language Bridging for Persian-Arabic and Hebrew-Arabic Statistical Machine Translation. NYUAD Annual Research Conference, 2015.
- [3] Sajjad, Hassan, Francisco Guzman, Preslav Nakov, Ahmed Abelali, Nadir Durrani, Wael Salloum, Ahmed El Kholy, and **Nizar Habash.** QCRI-Columbia-NYUAD submission to NIST Open MT Eval, 2015.

- [4] El-ballouli, Reem, Wassim El Hajj, Shady Elbassuoni, Hazem Hajj, **Nizar Habash**, Khaled Bashir Shaban. Credibility Models For Arabic Content On Twitter. Poster. Qatar Foundation Annual Research Conference, Doha, Qatar, 2014.
- [5] Baly, Ramy, Gilbert Badaro, Hazem Hajj, **Nizar Habash**, Wassim El Hajj, Khaled Shaban. Semantic Model Representation For Human's Pre-conceived Notions In Arabic Text With Applications To Sentiment Mining. Poster. Qatar Foundation Annual Research Conference, Doha, Qatar, 2014.
- [6] Zaghouani, Wajdi, **Nizar Habash**, Behrang Mohit, Abeer Heider, Alla Rozovskaya, Kemal Oflazer. Annotation Guidelines For Non-native Arabic Text In The Qatar Arabic Language Bank. Poster. Qatar Foundation Annual Research Conference, Doha, Qatar, 2014.
- [7] Badaro, Gilbert, Ramy Baly, Hazem Hajj, **Nizar Habash**, Wassim El-hajj, Khaled Shaban. An Efficient Model For Sentiment Classification Of Arabic Tweets On Mobiles. Poster. Qatar Foundation Annual Research Conference, Doha, Qatar, 2014.
- [8] Rozovskaya, Alla, **Nizar Habash**, Ramy Eskander, Noura Farra and Wael Salloum. The Columbia System in the QALB-2014 Shared Task on Arabic Error Correction. Arabic Natural Language Processing Workshop, EMNLP, Doha, 2014.
- [9] Rozovskaya, Alla, Kai-Wei Chang, Mark Sammons, Dan Roth and **Nizar Habash**. The Illinois-Columbia System in the CoNLL-2014 Shared Task, Baltimore, Maryland, 2014.
- [10] **Habash, Nizar**, Spence Green, Preslav Nakov, Shuly Wintner. Modeling Inflectional Morphology in Statistical MT targeting Morphologically Rich Languages. (**Group leader**). In Dagstuhl Seminar on "Statistical Techniques for Translating to Morphologically Rich Languages," Dagstuhl, Germany, 2014.
- [11] **Habash, Nizar**, Behrang Mohit, Ossama Obeid, Kemal Oflazer, Nadi Tomeh and Wajdi Zaghouani. QALB: Qatar Arabic Language Bank. Poster. Qatar Foundation Annual Research Conference, Doha, Qatar, 2013.
- [12] Marton, Yuval, **Nizar Habash**, Owen Rambow and Sarah Alkuhlani. SPMRL'13 Shared Task System: The CADIM Arabic Dependency Parser. 2013.
- [13] El Kholly, Ahmed, Wael Salloum and **Nizar Habash**. Columbia University 2012 Arabic-English, Farsi-English and Dari-English MT Evaluation Systems. NIST Machine Translation Evaluation, 2012.
- [14] Carpuat, Marine, Mahmoud Ghoneim, **Nizar Habash** and Mona Diab. Columbia University 2009 Arabic-English MT Evaluation System. NIST Machine Translation Evaluation, 2009.
- [15] Diab, Mona, **Nizar Habash** and Owen Rambow. Tools for Arabic Morphological Analysis, Disambiguation and Tokenization. Foreign Language Technology Forum, July 2008.
- [16] **Habash, Nizar**. Techniques for Online OOV Handling in Phrase-based Statistical Machine Translation. *Columbia University NLP Colloquium*, April 2008.
- [17] **Habash, Nizar**. Five Techniques for OOV Reduction in Arabic-English and Urdu-English SMT. *NIST MTEval Workshop*, March 2008.
- [18] **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.
- [19] **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.
- [20] 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.
- [21] **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.
- [22] 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 3.2)
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)
[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>]
- [5] CALIMA 0.4.2: Columbia Arabic Language Morphological Analyzer -- Egyptian Arabic (2012)
[Available from the Linguistic Data Consortium]
- [6] ELISSA 1.0: Dialectal Arabic to Modern Standard Arabic Translation System (2012, developed with Wael Salloum)
[Available from the Linguistic Data Consortium]
- [7] MADA-ARZ 0.3: Morphological Analysis and Disambiguation for Arabic (Egyptian version) (2012, developed with Owen Rambow and Ryan Roth)
[Available from the Linguistic Data Consortium]
- [8] Columbia Arabic Treebank Dependency Parser (2013, in collaboration with Yuval Marton, Owen Rambow, Ryan Roth and Sarah Alkuhlani)
[<http://www1.ccls.columbia.edu/~ymarton/pub/CATiB-parser/catib-parser.htm>]
- [9] DIRA: Dialectal Arabic Information Retrieval Assistant (2013, as part of the Columbia NADIA team; DIRA effort led by Nizar Habash and Owen Rambow)
Demo: <http://nlp.ldeo.columbia.edu/dira>
- [10] AIDA: Automatic Identification of Dialectal Arabic (2013, as part of the Columbia NADIA team; AIDA effort led by Mona Diab)
Demo: <http://nlp.ldeo.columbia.edu/aida>
- [11] MADAMIRA: A system for morphological analysis and disambiguation, base phrase chunking and named entity recognition. (2013, as part of the Columbia NADIA team; MADAMIRA merges the popular MADA tool [Habash and Rambow] and AMIRA tool [Diab])
Demo: <http://nlp.ldeo.columbia.edu/madamira> (English interface)
<http://nlp.ldeo.columbia.edu/madamira/#locale=ar> (Arabic interface)
Download: ([click here](#))
- [12] 3ARRIB: A system for transliteration from Romanized Arabic to Arabic script (2014, developed with Mohamed Al-Badrashiny, Ramy Eskander, and Owen Rambow)

Resources

- [13] Gumar: The Gulf Arabic Corpus (2016)
[<http://camel.abudhabi.nyu.edu/gumar/>]
- [14] ArSenL 1.0: Arabic Sentiment Lexicon (2014, developed with Gilbert Badaro, Ramy Baly, Hazem Hajj and Wassim El-Hajj).
[<http://oma-project.com/ArSenL/intro>]
- [15] QALB: Qatar Arabic Language Bank (2014, developed with Behrang Mohit, Wajdi Zaghouani, Ossama Obeid and Alla Rozovskaya)
Register [here to get a copy of QALB](#)

- [16] [Columbia Standard Arabic Gender-Number-Rationality Database](#) (2013, developed with Sarah Alkuhlani)
- [17] CATiB v1.0: Columbia Arabic Treebank (2009) [<http://www1.ccls.columbia.edu/~CATiB/>]
- [18] SEPIA: Machine Translation Evaluation Metric (2008) (developed with Ahmed El Kholly) [<http://www1.ccls.columbia.edu/~SEPIA/>]
- [19] CatVar A Categorical Variation Database for English (2002) [<http://clipdemos.umiacs.umd.edu/catvar/>]

Teaching Experience

Classes Taught

- [1] Undergraduate Core Course. CADT-UH 1011: Words. New York University Abu Dhabi, Fall 2017. 15 students.
- [2] Undergraduate Computer Science Course. CS-AD 220: Natural Language Processing. New York University Abu Dhabi, Spring 2017. 11 students. [Syllabus: <https://sites.google.com/a/nyu.edu/nyuad-cs-ad-220-natural-language-processing-spring-2017/>]
- [3] Undergraduate Core Course. COREI-AD 58: Words. New York University Abu Dhabi, Fall 2016. 14 students. [Syllabus: <https://sites.google.com/a/nyu.edu/nyuad-words-fall2016/>]
- [4] Undergraduate Computer Science Course. CS-AD 220: Natural Language Processing. New York University Abu Dhabi, Spring 2016. 12 students. [Syllabus: <https://sites.google.com/a/nyu.edu/nyuad-cs-ad-220-natural-language-processing-spring-2016/>]
- [5] Undergraduate Core Course. COREI-AD 58: Words. New York University Abu Dhabi, Fall 2015. 14 students. [Syllabus: <https://sites.google.com/a/nyu.edu/nyuad-words-fall2015/>]
- [6] Undergraduate Computer Science Course. CS-AD 219: Topics in Computer Science: Natural Language Processing. New York University Abu Dhabi, Spring 2015. 3 students. [Syllabus: <https://sites.google.com/a/nyu.edu/nyuad-nlp-spring2015/>]
- [7] Instructor of graduate-level course. E6998 section 003: Machine Translation. Department of Computer Science, Columbia University, Spring 2013. 13 students. [Syllabus: <https://sites.google.com/site/comse6998machinetranslation/>]
- [8] Instructor of graduate-level course. E6998 section 003: Machine Translation. Department of Computer Science, Columbia University, Fall 2011. 9 students.
- [9] Instructor of graduate-level course. E6998 section 003: Machine Translation. Department of Computer Science, Columbia University, Fall 2010. 21 students.
- [10] 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 Prof. Hend Al-Khalifa of King Saud University.
- I created several popular tutorials. See the list of conference tutorials and invited tutorials below.

Doctoral Committees

- [1] Ramy Baly, American University of Beirut, 2014 (proposal), 2016 (defense)
- [2] Leena Lulu, United Arab Emirates University, 2016 (defense)
- [3] Gilbert Badaro, American University of Beirut, 2016 (proposal)
- [4] Nawar Halabi, University of South Hampton, 2016 (defense)
- [5] Claudia Borg, University of Malta, 2016 (defense)

- [6] Mohammad Salameh, University of Alberta, 2016 (defense)
- [7] Ahmed El Kholy, Columbia University, 2012 (candidacy), 2013 (proposal), 2015 (defense) [my advisee]
- [8] Walid Aransa, L'Université du Maine, 2015 (defense)
- [9] Iskandar Keskes, L'Institut de Recherche en Informatique de Toulouse, 2015 (defense)
- [10] Wael Salloum, Columbia University, 2014 (candidacy) [my advisee]
- [11] Weiyun Ma, Columbia University, 2012 (candidacy), 2013 (proposal), 2014 (defense)
- [12] Kais Dukes, Leeds University, 2013 (defense) [secondary advisor]
- [13] Sarah Alkuhlani, Columbia University, 2013 (candidacy) [my advisee]
- [14] Kristen Parton, Columbia University, 2009 (candidacy), 2010 (proposal), 2012 (defense).
- [15] Sara Stymne, Linköping University, 2012 (defense; served as opponent)
- [16] Fadi Biadisy, Columbia University, 2008 (candidacy), 2010 (proposal), 2011 (defense).
- [17] Jakob Elming, Copenhagen Business School, 2009 (defense).
- [18] Josep Crego, Universitat Politècnica de Catalunya, 2008 (defense).
- [19] Otakar Smrz, Charles University, Czech Republic, 2007 (defense).
- [20] Andrew Rosenberg, Columbia University, 2006 (candidacy).

Masters Committees

- [1] Diyam Akra, Birzeit University, Computer Science, 2105 (defense)
- [2] Tejaswini Pepadati, Columbia University, Computer Science, 2014 (defense; external)
- [3] Ibrahim Saleh, Georgetown University, Linguistics, 2009 (Reader)

Undergraduate Academic Mentoring

- [1] Daniel Watson, NYUAD (Class of 2020)
- [2] Anh Le Xuan Mai, NYUAD (Class of 2020)
- [3] Gerardo Samuel Ixcaragua, NYUAD (Class of 2020)
- [4] Sangjin Lee, NYUAD (Class of 2020)
- [5] Tomasz Bachosz, NYUAD (Class of 2019)
- [6] Yana Chala, NYUAD (Class of 2019)
- [7] Guyu Fan, NYUAD (Class of 2019)
- [8] Theodore Tenev, NYUAD (Class of 2019)
- [9] Paula Valentina Dozsa, NYUAD (Class of 2019)
- [10] Muaz Ahmad, NYUAD (Class of 2019)
- [11] Dana Abu Ali, NYUAD (Class of 2018)
- [12] Angela Orozco, NYUAD (Class of 2018) – before she switched to Economics
- [13] Asyrique Asyraf Thevendran, NYUAD (Class of 2018)
- [14] Lingliang Zhang, NYUAD (Class of 2017)

Teaching Service

- [1] Ph.D. Program coordinator of the Computer Science Program, NYUAD, 2016-present.
- [2] Ph.D. student coordinator for the Center for Computational Learning Systems (CCLS) 2012-2014.

Guest Lectures

- [1] **Habash, Nizar.** The Palisra Project, COREA-AD 50W: Identity and Object, NYU Abu Dhabi, (Fall 2015, Fall 2016).

- [2] **Habash, Nizar.** On Language Construction, COREA-AD 19: Communication and Technology, NYU Abu Dhabi, (Fall 2015, Fall 2016).
- [3] **Habash, Nizar.** Stories, ARABL-AD 218: Colloquial Arabic: Levantine Dialect 1, NYU Abu Dhabi, (Fall 2015, Fall 2016).
- [4] **Habash, Nizar.** Introduction to Machine Translation, COMS 4701: Artificial Intelligence, Columbia University, **Fall 2014 (December 4, 2014).**
- [5] **Habash, Nizar.** Introduction to Arabic Natural Language Processing, CS 4705: Introduction to Natural Language Processing, Columbia University, **Fall 2013, Fall 2012, Fall 2011.**
- [6] **Habash, Nizar.** Machine Translation. Mock Class, New York University Abu Dhabi, **Spring 2013.**
- [7] **Habash, Nizar.** Machine Translation: Challenges, Approaches and Evaluation, CS 4705: Introduction to Natural Language Processing, Columbia University, **Fall 2010, Fall 2008-2004.**
- [8] **Habash, Nizar.** Machine Translation: Challenges and Approaches, CMSC 421: Introduction to Artificial Intelligence, University of Maryland College Park, **Fall 2003.**
- [9] **Habash, Nizar.** Generation Heavy Hybrid Machine Translation, CMSC 723/LING 645: Introduction to Computational Linguistics, University of Maryland College Park, **Spring 2003.**
- [10] **Habash, Nizar.** Computers and Writing Systems, HONR 279I: History of the Alphabets: Languages and their Scripts, University of Maryland College Park, **Spring 2003.**
- [11] **Habash, Nizar.** Natural Language Generation, CMSC 828: Advanced Natural Language Processing: Theory and Practice, University of Maryland College Park, **Spring 2002.**

Short Courses

- [1] **Habash, Nizar** and Owen Rambow. Princess Sumaya University of Technology (PSUT) Winter School on Arabic Language Processing, Amman, Jordan, January 27-29, 2013.
 - A three-day short course on Arabic NLP including talks, demos and discussion of students' research at PSUT.
 - Talks by Nizar Habash:
 - a. Introduction to Arabic Natural Language Processing
 - b. Processing Arabic Dialects
 - c. Morphological Analysis and Disambiguation of Arabic (talk and demo)
 - d. Machine Translation from and to Arabic

Conference Tutorials

- [1] Diab, Mona and **Nizar Habash.** Natural Language Processing of Arabic and its Dialects.
 - A three-hour tutorial. In the conference of Empirical Methods for Natural Language Processing (EMNLP), Doha, Qatar, 2014. 41 attendees.
- [2] **Habash, Nizar.** Machine Translation and Arabic Language Issues.
 - A three-hour tutorial. In the Association for Machine Translation in the Americas (AMTA'12), San Diego, CA, 2012.
- [3] Diab, Mona and **Nizar Habash.** Arabic Dialect Processing.
 - A three-hour tutorial. In the Conference of the North American chapter of the Association for Computational Linguistics (NAACL), Montreal, Canada, 2012.
 - 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.
- [4] **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.
- [5] **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.

See also *Invited Tutorials* under *Oral Presentations*

Service

Research Proposal Reviews

- [1] Review panel member, American Association for the Advancement of Science (AAAS) (2010, 2011, 2012, 2013, 2014).
- [2] Qatar National Research Fund reviewer (2009, 2010).
- [3] NSF review panel (2009, 2015).
- [4] NSF proposal reviewer (2005, 2006).

Journal Editor/Editorial Boards

- [1] Co-editor for a special issue on Arabic Language Processing as part of *la revue internationale Traitement Automatique des Langues* - <http://tal-58-3.sciencesconf.org/> (2017)
- [2] Member of the Editorial Board of Computer Speech and Language (2016-2017)
- [3] Member of the standing review committee of Transactions of the Association for Computational Linguistics (TACL) (2014-2016)
- [4] Member of the editorial board of Computational Linguistics Journal (2012-2014)
- [5] Co-editor with Hany Hassan for a special issue of MT journal on Arabic Machine Translation, 2012.
- [6] Member of the editorial board of the Journal for Machine Translation (2007-2015)

Journal Reviewing

- [1] Journal of Information Processing and Management (2016, 2017)
- [2] Journal of Knowledge-Based Systems (2016)
- [3] Journal of Computer Speech & Language (2016)
- [4] Journal of King Saud University (2016)
- [5] Language Resources and Evaluation Journal, (2015, 2014)
- [6] Foundations and Trends in IR, (2014)
- [7] Journal of Natural Language Engineering, (2004, 2007, 2011-2014).
- [8] Machine Translation Journal (2007-2014)
- [9] Transactions of the Association for Computational Linguistics (TACL) (2015, 2013).
- [10] Journal of Computational Linguistics, (2003, 2006, 2010-2014).
- [11] Journal of Computer Speech and Language (2013-2014)
- [12] Perspectives on Arabic Linguistics (2013, 2014)

- [13] Journal of Pattern Recognition Research (2011).
- [14] Journal of Artificial Intelligence Research (2011).
- [15] ACM Transaction on Asian Languages Information Processing (TALIP), (2004, 2010, 2014).
- [16] IEEE Transactions on Audio, Speech and Language Processing (2004, 2005, 2006, 2007).

Book Reviewing

- [1] A member of the scientific review committee for “Introduction to Computational Linguistics”, Rashwan and Elsaid, eds. (in Arabic). 2014.

Conference/Workshop Program Committees

- [1] International Conference of Arab Women in Computing: 2017
- [2] Universal Dependencies Workshop: 2017
- [3] Arabic Computational Linguistics Conference: 2017
- [4] Conference of the European ACL (EACL): 2017, 2016, 2014, 2012, 2006
- [5] Workshop on NLP for Similar Languages, Varieties and Dialects: 2017, 2016
- [6] Workshop on Semitic Machine Translation (SeMat): 2016
- [7] Workshop of the Special Interest Group on Computational Morphology and Phonology (SIGMORPHON): 2016
- [8] Workshop on Statistical Machine Translation (WMT): 2017, 2016, 2015, 2011, 2010
- [9] Conference of the Forum for Arabic Linguistics: 2015
- [10] Linguistic Annotations Workshop (LAW): 2017, 2015.
- [11] Conference of the Association for the Advancement of Artificial Intelligence (AAAI): 2014
- [12] Workshop on Free/Open-Source Arabic Corpora and Corpora Processing (at LREC): 2014, 2016
- [13] Workshop on Arabic Natural Language Processing: Models, Systems and Applications (at IEEE CIST'14): 2014
- [14] Palestinian Symposium on Computational Linguistics and Arabic Content (iArabic'14): 2014
- [15] International Conference on Parsing Technologies (IWPT): 2013
- [16] Conference on Lexical and Computational Semantics (*SEM): 2013
- [17] Conference on Computational Natural Language Learning (CONLL): 2013
- [18] Machine Translation Summit (MT Summit): 2013, 2011, 2007
- [19] Conference of the Association for Computational Linguistics (ACL): 2013, 2012, 2008-2006, 2002
- [20] Conference of the Association for Machine Translation in the Americas (AMTA): 2014-2006
- [21] Conference of the North American ACL (NAACL): 2015, 2013, 2012, 2007, 2006, 2001
- [22] Conference of the International Speech Communication Association (Interspeech): 2012
- [23] Conference on Computational Linguistics (COLING): 2008, 2004,
- [24] Conference on Empirical Methods for Natural Language Processing (EMNLP): 2107, 2009-2007
- [25] Conference on Mediterranean Arabic Language and Speech Technology (MEDAR): 2009
- [26] International Conference Arabic Language Technology (ALTIC): 2011
- [27] International Conference on the Arabic Language Processing (CITALA): 2007
- [28] International Natural Language Generation Conference (INLG): 2002
- [29] Language Resources and Evaluation Conference (LREC): 2014, 2012, 2008, 2006
- [30] Theoretical and Methodological Issues in Machine Translation (TMI): 2007
- [31] Workshop on Computational Approaches to Arabic Script Languages: 2009, 2004
- [32] Workshop on Computational Approaches to Semitic Languages: 2007, 2005

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) (2008-2014)
- [3] Research community representative in the board of the Association for Machine Translation in the Americas (AMTA) (2006-2008)
- [4] Secretary of the ACL Special Interest Group on Semitic Languages (2006-2009)

Conference and Workshop Organization

- [1] General Chair of the Workshop for Arabic Natural Language Processing (2017).
- [2] Co-organizer of the Workshop on Arabic Dialect Technologies (WARDAT), Abu Dhabi, 2016
- [3] SEM Eval Steering Committee (2016)
- [4] Co-chair of the Arabic Natural Language Processing Workshop in EMNLP (2014 and 2015). And a member of the team organizing shared task component of the workshop.
- [5] Publication co-chair for NAACL 2012.
- [6] Workshop program co-chair for NAACL 2009.
- [7] Senior Program Committee member for NAACL 2007 (machine translation track).
- [8] Program co-chair for the conference of the Association for Machine Translation in the Americas (AMTA'06)
- [9] One of three organizers for a workshop on Computational Approaches to Semitic Languages ACL 2005.
- [10] Chair of the Workshop on Determining Interlingua Utility for Machine Translation. AMTA-2004, Washington DC, 2004.
- [11] 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 Open Evaluation program, 2007.
- [4] Member of the machine translation team under the NSF funded MINDS workshop, 2006.

University Service

New York University Abu Dhabi

- [1] Judge in the 2017 Annual New York University Abu Dhabi (NYUAD) International Hackathon for the Social Good in the Arab World. (Spring 2017).
- [2] Chair of two three-year tenure-track faculty review committee (2017).
- [3] Ph.D. Program coordinator of the Computer Science Program, NYUAD, 2016-present.
- [4] Co-chair of the CS faculty search committee (for the year 2016-2017).
- [5] MC for the NYUAD Computer Science Open House (Fall 2016).
- [6] A member of the Core Curriculum Committee, NYUAD, Sep 1, 2016 - present.
- [7] Chair of a three-year tenure-track faculty review committee (2016).
- [8] A member of a Humanities Division faculty search committee on Digital Humanities (2016, 2017).
- [9] Mentor. Hackathon for Social Good in the Arab World. NYUAD, 2015 and 2016.
- [10] Tower of Babel Student Interest Group advisor (2015-present).
- [11] Co-director of the CS Unix Lab, 2015-2016.
- [12] A member of the Freshman Colloquium Committee, 2015-2016.
- [13] A member of the Institutional Review Board, NYUAD, April 15, 2015-Aug 31, 2016.
- [14] A member of the doctoral student selection committee, Computer Science, NYUAD January 20 2015-now.
- [15] A member of the CS faculty search committee (for the year 2014-2015)
- [16] Led discussion group of Freshmen Class summer reading as part of Marhaba Week:

- "Parable of the Sower" - 2015, NYUAD.
 "Bamboo Stalk" - 2016, NYUAD
- [17] Gave three talks on Machine Translation as part of the undergraduate recruitment effort at Candidate Weekend 2015, NYUAD, 2015.
 - [18] Judge. Sheikh Mohamed bin Zayed Scholars Program Selections, NYUAD, 2015.
 - [19] Judge. AngelHack at NYUAD, 2015.

Columbia University

- [20] Center for Computational Learning System – a member of an ad hoc faculty promotion committee, 2014.
- [21] Center for Computational Learning System – lead on the finance committee, 2013-2014.
- [22] Center for Computational Learning System – a member of the faculty search committee, 2013.

Honors

- [1] Mentored the team that was selected to represent NYUAD for the international Hult Prize competition (2016).
- [2] Mentored the *Arabic Snippets (Qusasat)* team which won the First Place Award and Audience Choice Award in the 2016 NYUAD Hackathon for Social Good in the Arab World. April 17, 2016.
- [3] Best Paper Award from the European Association for Machine Translation (EAMT) conference, Trento, Italy, 2012. The paper: Parton, Kristen, Nizar Habash, Kathleen McKeown, Gonzalo Iglesias, Adrià de Gispert. Can Automatic Post-Editing Make MT More Meaningful?
- [4] Received three Google Research Awards (in 2009, 2010 and 2012).
- [5] Best Paper Award from the Traitement Automatique des Langues Naturelles (TALN) conference, Montréal, Canada, 2010. The paper: Carpuat, Marine, Yuval Marton and Nizar Habash. Reordering Matrix Post-verbal Subjects for Arabic-to-English SMT.
- [6] Graduate Student Service Award, University of Maryland College Park, 2003.
- [7] The Samuel N. Alexander Fellowship, Association for Computing Machinery, Washington D.C. Chapter, 2001.
- [8] The Phi Kappa Phi Award of Excellence, Phi Kappa Phi, 1997/1998.
- [9] 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.
- [10] The College of Engineering and Technology Outstanding Scholar Award, Old Dominion University, 1997.
- [11] The Faculty Award in Computer Engineering, College of Engineering and Technology, Old Dominion University, 1997.
- [12] The Outstanding Individualized Study Student Award, College of Arts and Letters, Old Dominion University, 1997.
- [13] The Award of Academic Excellence, Academic Honors Program Old Dominion University, 1997.
- [14] The Meghan O'Connor Award for Academic Achievement and Community Service, Academic Honors Program, Old Dominion University, 1997.
- [15] Undergraduate Research Award, Academic Honors Program, Old Dominion University, 1994.

References

Available upon request