

# 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](mailto: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. 2020 to <i>present</i> .	<b>New York University Abu Dhabi, Computer Science, Full Professor.</b>
Sep. 2015 to present.	<b>New York University, Computer Science, Global Network Professor.</b>
Sep. 2019 to <i>Aug 2021</i> .	New York University Abu Dhabi, Computer Science, Head of Program.
Sep. 2014 to Aug 2020.	New York University Abu Dhabi, Computer Science, Associate Professor.
Feb. 2009 to May 2018.	Columbia University, Computer Science, Adjunct Associate Professor.
Jul. 2014 to Jun. 2016.	Birzeit University, Sina Institute, Adjunct Associate Professor.
Feb. 2009 to Aug. 2014.	Columbia University, Center for Computational Learning Systems, Research Scientist.
Jul. 2005 to Jan. 2009.	Columbia University, Center for Computational Learning Systems, Associate Research Scientist.
Jul. 2004 to Jun. 2005.	Columbia University, Post-doctoral Researcher.
Jul. 2003 to Jun. 2004.	University of Maryland College Park, Post-doctoral Researcher.
Jan. 1998 to May 2003.	University of Maryland College Park, Graduate Research Assistant.
Jan. 1999 to Jan. 2000.	Nuun Labs, Inc., Chief Technical Officer.
Jun. 1998 to Aug. 1998.	Cycorp, Inc., Intern Ontological Engineer.

## Research Experience

---

### Research Areas

- Primary research area is natural language processing (NLP) / computational linguistics (CL).
- Specialty sub-areas are Arabic NLP, computational morphology and machine translation.
- I founded the [Computational Approaches to Modeling Language \(CAMEL\) Lab](#) in New York University Abu Dhabi in 2015.
- I co-founded the Columbia Arabic Dialect Modeling (CADIM) group in 2005 with Owen Rambow and Mona Diab.

### Citation Metrics

- Citations 13,079; h-index 58; i10-index 201. [Google Scholar](#) (June 16, 2022).

## Past and Current Research Support

- [1] "Prediction and preview effects in Arabic word recognition." New York University Abu Dhabi Research Enhancement Fund. David Melcher (PI), Nizar Habash (co-PI), Diogo Almeida (co-PI), Laila Familiar. Awarded \$250,000. Two-year project. 2022-2024.
- [2] "ZAEBUC Phase 2: Longitudinal and spoken extensions of the Zayed Arabic-English Bilingual Undergraduate Corpus." Zayed University Research Incentive Fund (RIF). David Palfreyman (PI), Co-Investigators: Nizar Habash, Suha Karaki, Ahmad Aljanadbah. Awarded 300,000 AED. 2022-2023. RIF Award #R21072.
- [3] "A Multi-Genre Syntactic Treebank for Arabic." Nizar Habash (PI). New York University Abu Dhabi Research Enhancement Fund. Awarded \$40,000. 2019-2020.
- [4] "Arabic-English bilingual corpus of student writing." Zayed University Research Incentive Fund (RIF). David Palfreyman (PI), Co-Investigators: Afaf Al-Bataineh, Nizar Habash, Eric Atwell, Suha Karaki, Robyn Albers. Awarded 300,000 AED. 2019-2020. RIF Award #R19068.
- [5] "Center for Language Technologies and Digital Humanities." -- *Center Proposal Grant*. Nizar Habash (co-PI) and David Wisley (co-PI). New York University Abu Dhabi Research Enhancement Fund. Awarded \$40,000. 2018-2019.
- [6] "Morphological Annotation of Gulf Arabic." Nizar Habash (PI). New York University Abu Dhabi Research Enhancement Fund. Awarded \$40,000. 2016-2018.
- [7] "SAMER: Simplification of Arabic Masterpieces for Extensive Reading." Muhamed Al-Khalil (PI), Nizar Habash (co-PI) and Dris Soulimani (co-PI). New York University Abu Dhabi Research Enhancement Fund. Awarded \$181,000. Two-year project. 2016-2018.
- [8] "Workshop on Arabic Dialect Technologies." Nizar Habash (PI). New York University Abu Dhabi Institute Funding. \$50,600. 2016.
- [9] "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 (2/2016-1/2019).
- [10] "A Multilingual-to-Arabic Machine Translation Evaluation Dataset." Nizar Habash (PI). New York University Abu Dhabi Research Enhancement Fund. Awarded \$40,000. 2015.
- [11] "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 (11/2013- 10/2016).
- [12] "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.
- [13] "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.
- [14] "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 (4/2012-8/2015).
- [15] "Morphology Modeling for LORELEI." Owen Rambow (PI) and Nizar Habash. IBM Subcontract. IARPA BABEL Program. Approved. \$307,500 per year for four years (3/2012-6/2016).
- [16] "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).
- [17] "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).
- [18] "Preserving Morphological Richness in Poor-Pivot-based Statistical Machine Translation." Nizar Habash (PI). Google Research Award. 9/2012-8/2013 (\$84,211).
- [19] "Improving Statistical Machine Translation into Arabic." Nizar Habash (PI). Science Applications International Corporation (SAIC) research grant. 1/2012-8/2013 (\$120,307).
- [20] "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).

- [21] "Arabic Dialect Tools and Resources for COLABA." Mona Diab (PI), Nizar Habash and Owen Rambow. Axiom Subcontract. TSWG COLABA Program. P1 10/2008-3/2010 (\$442,709); P2 4/2010-3/2011(\$257,618).
- [22] "Arabic Dialect Translation for Agile." Nizar Habash (PI). IBM Subcontract. DARPA GALE Program. GALE-Y5 6/2010-6/2011 (\$112,437).
- [23] "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).
- [24] "Columbia Arabic Tree Bank." Nizar Habash (PI). IBM Subcontract. DARPA GALE Program. GALE-Y3 4/2008-8/2008 (\$200,000).
- [25] "SGER: Automatic Processing of Natural Language Code Switching." Mona Diab (PI), Nizar Habash and Owen Rambow. NSF. 9/2007-2/2009 (\$40,467).
- [26] "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).
- [27] "CINDOR: Conceptual Interlingua Document Retrieval." Owen Rambow (PI) and Nizar Habash (Post-doc). Textwise, Inc. 2/2005-5/2006 (\$38,534).
- [28] "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).
- [29] "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] Bashar Alhafni, Computer Science, New York University. 2020-
- [2] Alberto Chierici, Computer Science, New York University. 2018-
- [3] Nasser Zalmout, Computer Science, New York University. 2015-2019. Currently, *Applied Scientist* at Amazon.
- [4] Wael Salloum, Computer Science Department, Columbia University. 2010-2018. Currently, *Chief Science Officer* at Mendel.ai.
- [5] Ahmed El Kholi, Computer Science Department, Columbia University. 2008-2015. Currently, *Data Scientist* at Microsoft Research.

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

### ***Senior Colleagues***

- [6] Dr. Houda Bouamor, Assistant Professor, Carnegie Mellon University Qatar.
- [7] Dr. Wajdi Zaghouni, Lead Annotator, Carnegie Mellon University Qatar (with Kemal Oflazer, Behrang Mohit and Houda Bouamor on two projects). 2011-2018.
- [8] Dr. Jim Jing-Yan Wang, Research Scientist. Computational Approaches to Modeling Language Lab, New York University, Abu Dhabi, 2016.
- [9] Dr. Hieu Hoang, Research Scientist. Computational Approaches to Modeling Language Lab, New York University, Abu Dhabi, 2015.

### ***Post-doctoral***

- [10] Riad Belkebir, Post-doctoral researcher, New York University Abu Dhabi, 2020-2021.
- [11] Mai Oudah, Post-doctoral researcher, New York University Abu Dhabi, 2017-2019.
- [12] Alla Rozovskaya, Post-doctoral researcher, Columbia University Center for Computational Learning Systems. 2013-2014.
- [13] Thomas Lippincott, Post-doctoral researcher, Columbia University Center for Computational Learning Systems (with Owen Rambow). 2013-2014.

- [14] Nadi Tomeh, Post-doctoral researcher, Columbia University Center for Computational Learning Systems. (Primary research advisor). 2012-2013.
- [15] Abdelati Hawwari, Post-doctoral researcher, Columbia University Center for Computational Learning Systems (with Mona Diab and Owen Rambow). 2011-2012.
- [16] Marine Carpuat, Post-doctoral researcher, Columbia University Center for Computational Learning Systems (with Mona Diab and Owen Rambow). 2010-2011.
- [17] Yuval Marton, Post-doctoral researcher, Columbia University Center for Computational Learning Systems. (Primary research advisor). 2009-2011.
- [18] Ryan Roth, Post-doctoral researcher, Columbia University Center for Computational Learning Systems (with Mona Diab and Owen Rambow). 2008-2009.

### ***Doctoral***

- [19] Shahd Dibas, Linguistics, Philology and Phonetics, University of Oxford. 2020-2021.
- [20] Go Inoue, Computer Science, New York University. 2019-2021
- [21] Dima Taji, Computer Science, New York University. 2017-2019
- [22] Alexander Erdmann, Visiting Doctoral Student (Ohio State University, Boren Fellowship). Computational Approaches to Modeling Language Lab, New York University Abu Dhabi, 2016-2017. Research Assistant, 2017- 2020.
- [23] Hind Saddiki, Research Assistant. Co-supervised with Prof. Muhamed Al-Khalil. New York University Abu Dhabi, 2016 - present. Doctoral Student (Mohammed V University).
- [24] Sarah Alkuhlani, Computer Science Department, Columbia University. 2009-2015.
- [25] Mohammad Sadegh Rasooli, Ph.D. student, Computer Science Department, Columbia University. (Advisor: Nizar Habash 2012-2014; Michael Collins 2014-2018).
- [26] Noura Farra, Ph.D. student, Computer Science Department, Columbia University. (Advisor: Nizar Habash 2012-2014; Kathy McKeown 2014-graduation).
- [27] Kristen Parton, Ph.D. student, Computer Science Department, Columbia University. (Advisor: Kathy McKeown). 2010-2012.
- [28] Kais Dukes, Ph.D. student, University of Leeds. (Advisor: Eric Atwell; Secondary: Nizar Habash). 2009-2011.
- [29] Mohamed El-Tantawy, Ph.D. student, Computer Science Department, Columbia University. (Advisor: Owen Rambow). 2008-2011.
- [30] Fadi Biadisy, Ph.D. student, Computer Science Department, Columbia University. (Advisor: Julia Hirschberg). 2008-2011.
- [31] Josep M Crego. Ph.D. student, Computer Science Department, Universitat Politècnica de Catalunya. (Advisor: José Mariño). 2007-2009
- [32] Jakob Elming, Ph.D. student, Computer Science Department, Copenhagen Business School. (Advisor: Daniel Hardt). 2006-2009.
- [33] Mahmoud Ghoneim. Ph.D. student, Computer Science Department, Cairo University. (Columbia Advisor: Mona Diab). 2006-2009.
- [34] Necip Fazil Ayan, Ph.D. student. Computer Science Department, University of Maryland College Park. (Advisor: Bonnie Dorr). 2003-2005.

### ***Junior Colleagues***

- [35] Christian Khairallah, Research Assistant. Computational Approaches to Modeling Language Lab, New York University Abu Dhabi, 2021-present.
- [36] Injy Hamed, Research Assistant. Computational Approaches to Modeling Language Lab, New York University Abu Dhabi, 2021-present.
- [37] Muhammed Abu Odeh, Research Assistant. Computational Approaches to Modeling Language Lab, New York University Abu Dhabi, 2020-present.
- [38] Sabit Hasan, Research Assistant. Carnegie Mellon University Qatar (with Houda Bouamor), 2018-2019.
- [39] Fadhl Eryani, Research Assistant. Computational Approaches to Modeling Language Lab, New York University Abu Dhabi, 2017-2021.
- [40] Sara Hassan, Research Assistant. Computational Approaches to Modeling Language Lab, New York University Abu Dhabi, 2016-2017.

- [41] Maverick Alzate, Research Assistant. Computational Approaches to Modeling Language Lab, New York University Abu Dhabi, 2016.
- [42] Dima Taji, Research Assistant. Computational Approaches to Modeling Language Lab, New York University Abu Dhabi, 2015-2017.
- [43] Anas Shahrour, Research Assistant. Computational Approaches to Modeling Language Lab, New York University Abu Dhabi, 2015-2016.
- [44] Salam Khalifa, Research Assistant. Computational Approaches to Modeling Language Lab, New York University Abu Dhabi, 2015-present.
- [45] Ossama Obeid, Research Programmer, Carnegie Mellon University Qatar (with Behrang Mohit). 2011-2017. New York University Abu Dhabi, 2017-now.
- [46] Ryan Roth, Staff Associate, Columbia University Center for Computational Learning Systems (with Mona Diab and Owen Rambow). 2010-2013.
- [47] Ramy Eskander, Staff Associate, Columbia University Center for Computational Learning Systems (with Owen Rambow). 2012-2017.
- [48] Arfath Pasha, Staff Associate, Columbia University Center for Computational Learning Systems (with Owen Rambow and Mona Diab). 2012-2015.
- [49] Mohamed Badrashiny, Staff Associate, Columbia University Center for Computational Learning Systems (with Owen Rambow and Mona Diab). 2012-2013.
- [50] Sakhar ElKhereyf. Computer Research Institute, King Abdulaziz City for Science and Technology (KACST). Fall 2013. Independent Researcher.

### ***Masters***

- [51] Go Inoue, Visiting Masters Student (NARA Institute of Science and Technology, Tobitate Fellowship). Computational Approaches to Modeling Language Lab, New York University Abu Dhabi, 2017-2018.
- [52] Sara Alwaalan. M.S. student, Computer Science Department, Columbia University. Fall 2013.
- [53] Amit Abbi. M.S. student, Computer Science Department, Columbia University (with Owen Rambow and Mona Diab). 2010-2011.
- [54] Reshef Shilon, M.S. student, Tel Aviv University. (Advisor: Shuly Wintner). 2010-2011.
- [55] Jun Hu, M.S. student, Computer Science Department, Columbia University. (Primary research advisor). 2008-2009.
- [56] Ryan Roth, M.S. student, Computer Science Department, Columbia University (with Owen Rambow). 2007.
- [57] Abdelhalim Abbas, M.S. student, Computer Science Department, Columbia University. (Primary research advisor). 2006.
- [58] Farhad Peydaye Saheli, M.S. student, Computer Science Department, Columbia University (with Owen Rambow). 2006.
- [59] Fadi Biadsy, M.S. student, Computer Science Department, Ben-Gurion University. (Advisor: Jihad El-Sana). 2005-2006.
- [60] Matthew Chu, M.S. student, Computer Science Department, Columbia University (with Owen Rambow). 2005.

### ***Undergraduate Capstone Advising***

- [61] Ons Taktak, undergraduate student, Computer Science, NYUAD, 2022-2023.
- [62] Shamma Alkhoori, undergraduate student, Computer Science, NYUAD, 2022-2023.
- [63] Nouf Abduljalil Abbasi, undergraduate student, Computer Science, NYUAD, 2022-2023.
- [64] Ahmed Elshabrawy, undergraduate student, Computer Science, NYUAD, 2022-2023.
- [65] Gautham Dinesh Kumar, undergraduate student, Computer Science, NYUAD, 2022-2023.
- [66] Muhammad Ali, undergraduate student, Computer Science, NYUAD, 2022-2023.
- [67] Long Phan, undergraduate student, Computer Science, NYUAD, 2022-2023.
- [68] Omar Kallas, undergraduate student, Computer Science, NYUAD, 2022-2023.
- [69] Aizaz Ansari, undergraduate student, Computer Science, NYUAD, 2020-2021.
- [70] Filip Bendeviski, undergraduate student, Computer Science, NYUAD, 2020-2021.
- [71] Jumana Ibrahim, undergraduate student, Computer Science, NYUAD, 2020-2021.
- [72] Kertu Koss, undergraduate student, Computer Science, NYUAD, 2020-2021.
- [73] Muhammad Anas Jawed, undergraduate student, Computer Science, NYUAD, 2020-2021.

- [74] Nurpeiis Baimukan, undergraduate student, Computer Science, NYUAD, 2020-2021.
- [75] Tina Krulec, undergraduate student, Computer Science, NYUAD, 2020-2021.
- [76] Tyece Hensley, undergraduate student, Computer Science, NYUAD, 2020-2021.
- [77] Wahib Kamran, undergraduate student, Computer Science, NYUAD, 2020-2021.
- [78] Ali Shazal, undergraduate student, Computer Science, NYUAD, 2019-2020.
- [79] Aiza Usman, undergraduate student, Computer Science, NYUAD, 2019-2020.
- [80] Ross Jiang, undergraduate student, Computer Science, NYUAD, 2019-2020.
- [81] Ella Noll, undergraduate student, Computer Science, NYUAD, 2018-2019.
- [82] Christine Chung, undergraduate student, Computer Science, NYUAD, 2018-2019.
- [83] Ali Al Dhaheri, undergraduate student, Computer Science, NYUAD, 2018-2019.
- [84] William Held, undergraduate student, Computer Science, NYUAD, 2018-2019.
- [85] Dana Abu Ali, undergraduate student, Computer Science, NYUAD, 2016-2018.
- [86] Hayat Abual Hassan, undergraduate student, Computer Science, NYUAD, 2017-2018.
- [87] Jihyun Kim, undergraduate student, Computer Science, NYUAD, 2017-2018.
- [88] Ming Hu, undergraduate student, Computer Science, NYUAD, 2017-2018.
- [89] Muaz Ahmad, undergraduate student, Computer Science, NYUAD, 2017-2018.
- [90] Paula Dozsa, undergraduate student, Computer Science, NYUAD, 2017-2018.
- [91] Shehroze Khan, undergraduate student, Computer Science, NYUAD, 2017-2018.
- [92] Tarik Zulfikarpasic, undergraduate student, Computer Science, NYUAD, 2017-2018.
- [93] Wei Chen, undergraduate student, Computer Science, NYUAD, 2017-2018.
- [94] Jose Varias, undergraduate student, Computer Science, NYUAD, 2017-2017.
- [95] Talha Javed, undergraduate student, Computer Science, NYUAD, 2017.
- [96] Morgante Pell, undergraduate student, Computer Science, NYUAD, 2015-2016.
- [97] Chukwuyem Onyibe, undergraduate student, Computer Science, NYUAD, 2015-2017.
- [98] Lingliang Zhang, undergraduate student, Computer Science, NYUAD, 2015-2016.
- [99] Jermsak "Mick" Jermsurawong, undergraduate student, Computer Science, NYUAD, 2014-2015.

### ***Undergraduate Research***

- [100] Omar Kallas ,undergraduate research assistant, NYUAD, Spring 2021.
- [101] Reem Hazim ,undergraduate research assistant, NYUAD, Spring 2021.
- [102] Armaan Agrawal ,undergraduate research assistant, NYUAD, Spring 2021.
- [103] Al Reem Al Hosani, undergraduate research assistant, NYUAD, Fall 2018.
- [104] Daniel Watson, undergraduate research assistant, NYUAD, Summer 2017-2018.
- [105] Fatma Al Fardan, undergraduate research assistant, NYUAD, Fall 2017.
- [106] William Held, undergraduate student, Computer Science, NYUAD, Spring 2016-2017.
- [107] Arvind Srinivasan, undergraduate student, Computer Science Department, Columbia University. Fall 2013.
- [108] Matt Capetola, undergraduate student, Computer Science Department, Columbia University (with Owen Rambow). 2010-2011.
- [109] Jacob Andreas, undergraduate student, Computer Science Department, Columbia University (with Owen Rambow). 2010-2011.
- [110] Morgan Rhodes, undergraduate student, Computer Science Department, Columbia University. 2006-2007.
- [111] Amitai Aviram, undergraduate student, Computer Science Department, Columbia University (with Owen Rambow). 2005.

### ***High School***

- [112] 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] Salloum, Wael and **Nizar Habash**. "Unsupervised Arabic dialect segmentation for machine translation." *Natural Language Engineering*. Cambridge University Press. Volume 28, Issue 2, pp. 223 - 248. 2022. DOI:10.1017/S1351324920000455
- [2] Jensen, Jeffrey, Daniel Karell, Cole Tanigawa-Lau, **Nizar Habash**, Mai Oudah, Dhia Fairus shofia Fani. *Language Models in Sociological Research: An Application to Classifying Large Administrative Data and Measuring Religiosity*. *Sociological Methodology*. 2021. DOI:10.1177/00811750211053370
- [3] Darwish, Kareem, **Nizar Habash**, Mourad Abbas, Hend Al-Khalifa, Hussein T. Al-Natsheh, Houda Bouamor, Karim Bouzoubaa, Violetta Cavalli- Sforza, Samhaa R. El-Beltagy, Wassim El-Hajj, Mustafa Jarrar, and Hamdy Mubarak. "A Panoramic Survey of Natural Language Processing in the Arab World". Special Issue on the Arab World. *Communications of the ACM* 64, no. 4 (2021) : 72-81.
- [4] Badaro, Gilbert, Hazem Hajj, and **Nizar Habash**. "A Link Prediction Approach for Accurately Mapping a Large-scale Arabic Lexical Resource to English WordNet." *ACM Transactions on Asian and Low-Resource Language Information Processing (TALLIP)* 19, no. 6 (2020): 1-38.
- [5] Badaro, Gilbert, Ramy Baly, Hazem Hajj, Wassim El-Hajj, Khaled Shaban, Nizar Habash, Ahmad Sallab, and Ali Hamdi. A Survey of Opinion Mining in Arabic: A Comprehensive System Perspective Covering Challenges and Advances in Tools, Resources, Models, Applications and Visualizations. In *ACM Transactions on Asian Language Information Processing (TALLIP)*. 2019.
- [6] Baly, Ramy, Hazem Hajj, **Nizar Habash**, Khaled Bashir Shaaban, and Wassim El-Hajj. A Sentiment Treebank and Morphologically Enriched Recursive Deep Models for Effective Sentiment Analysis in Arabic. In *ACM Transactions on Asian and Low-Resource Language Information Processing (TALLIP)*, 16(4), 23, 2017.
- [7] 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.
- [8] 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.
- [9] 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.
- [10] 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*.
- [11] Kais Dukes, Eric Atwell and **Nizar Habash**. "Supervised Collaboration for Syntactic Annotation of Quranic Arabic", *Language Resource and Evaluation Journal*, 47:33-62, 2013.
- [12] 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.
- [13] 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.
- [14] 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.
- [15] 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.

- [16] **Habash, Nizar**, Bonnie J. Dorr and Chirstof Monz. Symbolic-to-Statistical Hybridization: Extending Generation-Heavy Machine Translation. *Machine Translation Journal*, 23:23--63, 2009.
- [17] **Habash, Nizar**, Bonnie Dorr, and David Traum. Hybrid Natural Language Generation from Lexical Conceptual Structures. *MT Journal* volume 18 (2): 81-128, 2003.
- [18] 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.
- [19] **Habash, Nizar**. "Nuun: A System for Developing Platform and Browser Independent Arabic Web Applications." (In Arabic) *Arab Journal of Science*, 33, 1999.

## Books

- [20] **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

- [21] **Nizar Habash** and David Palfreyman. "ZAEBUC: Design and Annotation: Guidelines, Processes, and Insights." In *Bilingual Writers and Corpus Analysis*. Routledge, in press, 2022.
- [22] **Habash, Nizar**. "Arabic Computational Linguistics." In Ryding, Karin and David Wilmsen, eds. "Handbook of Arabic Linguistics." Cambridge University Press. 2021.
- [23] **Habash, Nizar**. "Arabic Dialect Processing." In Zampieri, Marcos and Preslav Nakov, eds. "Similar Languages, Varieties, and Dialects: A Computational Perspective." Cambridge University Press. 2021.
- [24] **Habash, Nizar**. "Case Study: the MADAMIRA System for Arabic Morphological Analysis, Automatic Diacritization, and Lexical Disambiguation." In *The Arabic Language Status Report*. UAE Ministry of Culture and Youth. 2020. «دراسة حالة: نظام مداميرا للتحليل الصرفي والتشكيل الآلي وحل»  
الانتباس المعجمي» في تقرير حالة اللغة العربية ومستقبلها.
- [25] 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., "Natural Language Processing of Semitic Languages," Springer, 2014. Pages 3-41.
- [26] **Nizar Habash** and Fatiha Sadat. "Arabic Preprocessing for Statistical Machine Translation: Schemes, Techniques and Combinations." In Soudi, A., Neumann, G., Farghaly, A., Zbib, R. eds, "Challenges for Arabic Machine Translation", Chapter 5, John Benjamins NLP series, 2012.
- [27] 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.
- [28] 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.
- [29] 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.
- [30] **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.
- [31] **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.

- [32] 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.
- [33] 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.
- [34] **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.
- [35] **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

### 2022

- [36] Alhafni, Bashar, **Nizar Habash**, Houda Bouamor. User-Centric Gender Rewriting . In Proceedings of the North American Association for Computational Linguistics (NAACL 2022), Seattle, 2022.
- [37] Abdulrahim, Dana, Go Inoue, Latifa Shamsan, Salam Khalifa and **Nizar Habash**. The Bahrain Corpus: A Multi-genre Corpus of Bahraini Arabic. In Proceedings of the International Conference on Language Resources and Evaluation (LREC 2022), Marseille, France, 2022.
- [38] Alhafni, Bashar, **Nizar Habash** and Houda Bouamor. The Arabic Parallel Gender Corpus 2.0: Extensions and Analyses. In Proceedings of the International Conference on Language Resources and Evaluation (LREC 2022), Marseille, France, 2022.
- [39] Baimukan, Nurpeiis, Houda Bouamor and **Nizar Habash**. Hierarchical Aggregation of Dialectal Data for Arabic Dialect Identification. In Proceedings of the International Conference on Language Resources and Evaluation (LREC 2022), Marseille, France, 2022.
- [40] Batsuren, Khuyagbaatar, Omer Goldman, Salam Khalifa, **Nizar Habash**, Witold Kieraś, and others. UniMorph 4.0: Universal Morphology. In Proceedings of the International Conference on Language Resources and Evaluation (LREC 2022), Marseille, France, 2022.
- [41] **Habash, Nizar** and David Palfreyman. ZAEBUC: An Annotated Arabic-English Bilingual Writer Corpus. In Proceedings of the International Conference on Language Resources and Evaluation (LREC 2022), Marseille, France, 2022.
- [42] **Habash, Nizar**, Muhammed AbuOdeh, Dima Taji, Reem Faraj, Jamila El Gizuli and Omar Kallas. Camel Treebank: An Open Multi-genre Arabic Dependency Treebank. In Proceedings of the International Conference on Language Resources and Evaluation (LREC 2022), Marseille, France, 2022.
- [43] Inoue, Go, Salam Khalifa, and **Nizar Habash**. Morphosyntactic Tagging with Pre-trained Language Models for Arabic and its Dialects. In Findings of the Association for Computational Linguistics (ACL 2022), Dublin, 2022.

### 2021

- [44] Belkebir, Riadh, and **Nizar Habash**. Automatic Error Type Annotation for Arabic. In Proceedings of the Conference on Computational Natural Language Learning (CONLL), Virtual, 2021.
- [45] Bendeviski, Filip, Jumana Ibrahim, Tina Krulec, Theodore Waters, **Nizar Habash**, Hanan Salam, Himadri Mukherjee, Christin Camia. Towards Automatic Narrative Coherence Prediction. In Proceedings of the ACM International Conference on Multimodal Interaction (ICMI '21), Virtual, 2021.
- [46] Chierici, Alberto , Tyece Hensley, Wahib Kamran, Kertu Koss, Armaan Agrawal, Erin Collins, Goffredo Puccetti and **Nizar Habash**. A Cloud-based User-Centered Time-Offset

Interaction Application. In Proceedings of the Meeting of the Special Interest Group on Discourse and Dialogue, Virtual, 2021.

## 2020

- [47] Zalmout, Nasser and **Nizar Habash**. Utilizing Subword Entities in Character-Level Sequence-to-Sequence Lemmatization Models. In Proceedings of the International Conference on Computational Linguistics (COLING 2020), Online, 2020.
- [48] Kankanampati, Yash, Joseph Le Roux, Nadi Tomeh, Dima Taji, **Nizar Habash**. Multitask Easy-First Dependency Parsing: Exploiting Complementarities of Different Dependency Representations. In Proceedings of the International Conference on Computational Linguistics (COLING 2020), Online, 2020.
- [49] Jiang, Zhengyang, **Nizar Habash** and Muhamed Al Khalil. An Online Readability Leveled Arabic Thesaurus. In Proceedings of the International Conference on Computational Linguistics (COLING 2020), Online, 2020.
- [50] Erdmann, Alexander, Micha Elsner Shijie Wu, Ryan Cotterell, and **Nizar Habash**. The Paradigm Discovery Problem. In Proceedings of Conference of the Association for Computational Linguistics (ACL), Seattle, Washington, USA, 2020.
- [51] Zalmout, Nasser and **Nizar Habash**. Joint Diacritization, Lemmatization, Normalization, and Fine-Grained Morphological Tagging. In Proceedings of Conference of the Association for Computational Linguistics (ACL), Seattle, Washington, USA, 2020.
- [52] Al Khalil, Muhamed, **Nizar Habash**, Zhengyang Jiang. A Large-Scale Leveled Readability Lexicon for Standard Arabic. In Proceedings of the International Conference on Language Resources and Evaluation (LREC 2020), Marseille, France, 2020.
- [53] Chierici, Alberto M., **Nizar Habash**, Margarita Bicec. The Margarita Dialogue Corpus: A Data Set for Time-Offset Interactions and Unstructured Dialogue Systems. In Proceedings of the International Conference on Language Resources and Evaluation (LREC 2020), Marseille, France, 2020.
- [54] Eryani, Fadhl, **Nizar Habash**, Houda Bouamor, Salam Khalifa. A Spelling Correction Corpus for Multiple Arabic Dialects. In Proceedings of the International Conference on Language Resources and Evaluation (LREC 2020), Marseille, France, 2020.
- [55] Khalifa, Salam, Nasser Zalmout, **Nizar Habash**. Morphological Analysis and Disambiguation for Gulf Arabic: The Interplay between Resources and Methods. In Proceedings of the International Conference on Language Resources and Evaluation (LREC 2020), Marseille, France, 2020.
- [56] Obeid, Ossama, Nasser Zalmout, Salam Khalifa, Dima Taji, Mai Oudah, Bashar Alhafni, Go Inoue, Fadhl Eryani, Alexander Erdmann, **Nizar Habash**. CAMEL Tools: An Open Source Python Toolkit for Arabic Natural Language Processing. In Proceedings of the International Conference on Language Resources and Evaluation (LREC 2020), Marseille, France, 2020.

## 2019

- [57] Obeid, Ossama, Mohammad Salameh, Houda Bouamor and **Nizar Habash**. ADIDA: Automatic Dialect Identification for Arabic. In Proceedings of the Conference of the North American Association for Computational Linguistics (NAACL), Minneapolis, 2019.
- [58] Zalmout, Nasser and **Nizar Habash**. Adversarial Multitask Learning for Joint Multi-Feature and Multi-Dialect Morphological Modeling. In Proceedings of Conference of the Association for Computational Linguistics (ACL), Florence, Italy, 2019.
- [59] Held, William and **Nizar Habash**. The Effectiveness of Simple Hybrid Systems for Hypernym Discovery. In Proceedings of Conference of the Association for Computational Linguistics (ACL), Florence, Italy, 2019.
- [60] Oudah, Mai, Amjad Almahairi, and **Nizar Habash**. The Impact of Preprocessing on Arabic-English Statistical and Neural Machine Translation. In Proceedings of the Machine Translation Summit, Dublin, Ireland, 2019.
- [61] Ali, Ahmed, Salam Khalifa, and **Nizar Habash**. Towards Variability Resistant Dialectal Speech Evaluation. In Proceedings of Interspeech. Graz, Austria, 2019.

## 2018

- [62] Watson, Daniel, Nasser Zalmout and **Nizar Habash**. Utilizing Character and Word Embeddings for Text Normalization with Sequence-to-Sequence Models. In Proceedings of the Empirical Methods for Natural Language Processing Conference, Brussels, 2018.
- [63] Salameh, Mohammad, Houda, Bouamor, and **Nizar Habash**. Fine-Grained Arabic Dialect Identification. In Proceedings of the Computational Linguistics Conference (COLING), Santa Fe, New Mexico, 2018.
- [64] Khan, Shehroze, Jihyun Kim, Tarik Zulfikarpasic, Peter Chen and **Nizar Habash**. A Cross-lingual Messenger with Keyword Searchable Phrases for the Travel Domain. In Proceedings of the Computational Linguistics Conference (COLING), Santa Fe, New Mexico, 2018.
- [65] Abu Ali, Dana, Muaz Ahmad, Hayat Al Hassan, Paula Dozsa, Ming Hu, Jose Varias and **Nizar Habash**. A Bilingual Interactive Human Avatar Dialogue System. In Proceedings of the Conference of the Special Interest Group on Discourse and Dialogue (SIGDIAL), Melbourne, Australia, 2018.
- [66] Erdmann, Alexander, Nasser Zalmout and **Nizar Habash**. Addressing Noise in Multidialectal Word Embeddings. In Proceedings of Conference of the Association for Computational Linguistics (ACL), Melbourne, Australia, 2018.
- [67] Zalmout, Nasser, Alexander Erdmann, and **Nizar Habash**. Noise-Robust Morphological Disambiguation for Dialectal Arabic. In Proceedings of Conference of the North American Association for Computational Linguistics (NAACL), New Orleans, 2018.
- [68] **Habash, Nizar**, Fadhl Eryani, Salam Khalifa, Owen Rambow, Dana Abdulrahim, Alexander Erdmann, Reem Faraj, Wajdi Zaghouni, Houda Bouamor, Nasser Zalmout, Sara Hassan, Faisal Al Shargi, Sakhar Alkhereyf, Basma Abdulkareem, Ramy Eskander, Mohammad Salameh and Hind Saddiki. Unified Guidelines and Resources for Arabic Dialect Orthography. In Proceedings of the International Conference on Language Resources and Evaluation (LREC 2018), Miyazaki, Japan, 2018.
- [69] Javed, Talha, **Nizar Habash** and Dima Taji. Palmyra: A Platform Independent Dependency Annotation Tool for Morphologically Rich Languages. In Proceedings of the International Conference on Language Resources and Evaluation (LREC 2018), Miyazaki, Japan, 2018.
- [70] Obeid, Ossama, Salam Khalifa, **Nizar Habash**, Houda Bouamor, Wajdi Zaghouni and Kemal Oflazer. MADARi: A Web Interface for Joint Arabic Morphological Annotation and Spelling Correction. In Proceedings of the International Conference on Language Resources and Evaluation (LREC 2018), Miyazaki, Japan, 2018.
- [71] Bouamor, Houda, **Nizar Habash**, Mohammad Salameh, Wajdi Zaghouni, Owen Rambow, Dana Abdulrahim, Ossama Obeid, Salam Khalifa, Fadhl Eryani, Alexander Erdmann and Kemal Oflazer. The MADAR Arabic Dialect Corpus and Lexicon. In Proceedings of the International Conference on Language Resources and Evaluation (LREC 2018), Miyazaki, Japan, 2018.
- [72] Inoue, Go, **Nizar Habash**, Yuji Matsumoto and Hiroyuki Aoyama. A Parallel Corpus of Arabic-Japanese News Articles. In Proceedings of the International Conference on Language Resources and Evaluation (LREC 2018), Miyazaki, Japan, 2018.
- [73] Khalifa, Salam, **Nizar Habash**, Fadhl Eryani, Ossama Obeid, Dana Abdulrahim and Meera Al Kaabi. A Morphologically Annotated Corpus of Emirati Arabic. In Proceedings of the International Conference on Language Resources and Evaluation (LREC 2018), Miyazaki, Japan, 2018.
- [74] Al-Khalil, Muhamed, Hind Saddiki, **Nizar Habash**, and Latifa Alfalasi. A Leveled Reading Corpus of Modern Standard Arabic. In Proceedings of the International Conference on Language Resources and Evaluation (LREC 2018), Miyazaki, Japan, 2018.
- [75] More, Amir, Özlem Çetinoğlu, Çağrı Çöltekin, **Nizar Habash**, Benoît Sagot, Djamé Seddah, Reut Tsarfaty, and Dima Taji. CoNLL-UL: Universal Morphological Lattices for Universal Dependency Parsing. In Proceedings of the International Conference on Language Resources and Evaluation (LREC 2018), Miyazaki, Japan, 2018.

## 2017

- [76] Al-Khalil, Mohamed, **Nizar Habash** and Hind Saddiki. Simplification of Arabic Masterpieces for Extensive Reading: A Project Overview. In Proceedings of the International Conference on Arabic Computational Linguistics, Dubai, UAE, 2017.

- [77] Zalmout, Nasser, and **Nizar Habash**. Don't Throw Those Morphological Analyzers Away Just Yet: Neural Morphological Disambiguation for Arabic. In Proceedings of the conference on Empirical Methods for Natural Language Processing (EMNLP), Copenhagen, Denmark, 2017.
- [78] Erdmann, Alexander, **Nizar Habash**, Dima Taji and Houda Bouamor. Low Resourced Machine Translation via Morpho-syntactic Modeling: The Case of Dialectal Arabic. In Proceedings of the Machine Translation Summit, Nagoya, Japan, 2017.
- [79] **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.
- [80] Zalmout, Nasser and **Nizar Habash**. Optimizing Tokenization Choice for Machine Translation across Multiple Target Languages. In Proceedings of the European Association of Machine Translation (EAMT). *The Prague Bulletin of Mathematical Linguistics No. 108, (2017)*, pp. 257-269. doi: 10.1515/pralin-2017-0025. Prague, Czech Republic, 2017.
- [81] 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

- [82] 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.
- [83] 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.
- [84] 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.
- [85] 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.
- [86] 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.
- [87] 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.
- [88] 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.
- [89] Al-Shargi, Faisal, Aidan Kaplan, Ramy Eskander, **Nizar Habash** and Owen Rambow. Morphologically annotated corpora and morphological analyzers for Moroccan and Sanaani

- Yemeni Arabic. In Proceedings of the Language Resources and Evaluation Conference (LREC), Portorož, Slovenia, 2016.
- [90] **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.
- [91] 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.
- [92] 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.
- [93] 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.
- [94] 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

- [95] El Kholy, Ahmed, and **Nizar Habash**. Morphological Constraints for Phrase Pivot Statistical Machine Translation. In Proceedings of the Machine Translation Summit. Miami, 2015.
- [96] Shahrour, Anas, Salam Khalifa and **Nizar Habash**. Improving Arabic Diacritization through Syntactic Analysis. In Proceedings of EMNLP, Lisbon, 2015.
- [97] Jermsurawong, Jermsak and **Nizar Habash**. Predicting the structure of Cooking Recipes. In Proceedings of the conference on Empirical Methods for Natural Language Processing (EMNLP), Lisbon, 2015.
- [98] 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

- [99] Sikos, Jennifer, Peter David, **Nizar Habash**, and Reem Faraj. Authorship Analysis of Inspire Magazine (Jihadi Writing) through Stylometric and Psychological Features. Joint Intelligence and Security Informatics Conference (JISIC) Proceedings, September 2014.
- [100] 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.
- [101] El Kholy, 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.
- [102] 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.
- [103] 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.
- [104] 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.
- [105] 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.

- [106] Rasooli, Mohammad Sadegh, **Nizar Habash**, Owen Rambow and Thomas Lippincott. Unsupervised Morphology-Based Vocabulary Expansion. In Proceedings of the Conference of the Association for Computational Linguistics (ACL), Baltimore, Maryland, 2014.
- [107] 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.
- [108] 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.
- [109] 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.
- [110] Masmoudi, Abir, Mariem Ellouze Khmekhem, Yannick Esteve, Lamia Hadrich 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.
- [111] 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.
- [112] Zaghrouani, 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.
- [113] 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

- [114] 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.
- [115] Hamdi, Ahmed, Rahma Boujelbane, **Nizar Habash** and Alexis Nasr. The Effects of Factorizing Root and Pattern Mapping in Bidirectional Tunisian - Standard Arabic Machine Translation. In Proceedings of the Machine Translation Summit (MTSummit), Nice, France, 2013.
- [116] 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.
- [117] 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.
- [118] Obeid, Ossama, Wajdi Zaghrouani, 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.
- [119] 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.
- [120] 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.
- [121] 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.
- [122] **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.
- [123] 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.
- [124] 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.
- [125] 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.
- [126] 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 des Langues Naturelles 2013 (TALN 2013), Les Sables-d'Olonne, France, 2013.

## 2012

- [127] 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.
- [128] 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.
- [129] El Kholly, 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.
- [130] El Kholly, 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.
- [131] 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.
- [132] 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.
- [133] 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.
- [134] **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

- [135] El Kholly, Ahmed 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.

- [136] 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.
- [137] 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.
- [138] **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.
- [139] 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

- [140] 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.
- [141] Carpuat, Marine, Yuval Marton and **Nizar Habash**. Reordering Matrix Post-verbal Subjects for Arabic-to-English SMT. In Proceedings of Traitement Automatique des Langues Naturelles (TALN-10). Montréal, Canada, 2010.
- [142] Kholly, 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.
- [143] 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.
- [144] Dukes, Kais and **Nizar Habash**. Morphological Annotation of Quranic Arabic. In Proceedings of the Language Resources and Evaluation Conference (LREC), Malta, 2010.

## 2009

- [145] **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.
- [146] Biadisy, Fadi, **Nizar Habash** and Julia Hirschberg. Improving the Arabic Pronunciation Dictionary for Phone and Word Recognition with Linguistically Based Pronunciation Rules. In Proceedings of the North American Association for Computational Linguistics (NAACL), Boulder, 2009.
- [147] **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.
- [148] **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.
- [149] **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

- [150] **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.
- [151] **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.
- [152] 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.

[153] **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.

[154] 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

[155] **Habash, Nizar**. Syntactic Preprocessing for Statistical Machine Translation. In Proceedings of the Machine Translation Summit (MT-Summit), Copenhagen, Denmark, 2007.

[156] 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.

[157] 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.

[158] **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.

[159] **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.

[160] 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.

[161] **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

[162] **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.

[163] **Habash, Nizar** and Owen Rambow. MAGEAD: A Morphological Analyzer and Generator for the Arabic Dialects. In Proceedings of ACL, Sydney, Australia, 2006.

[164] Sadat, Fatiha and **Nizar Habash**. Combination of Arabic Preprocessing Schemes for Statistical Machine Translation. In Proceedings of ACL, Sydney, Australia, 2006.

[165] **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.

[166] Chiang, David, Mona Diab, **Nizar Habash**, Owen Rambow, and Safi Shareef. Parsing Arabic Dialects. In Proceedings of the European chapter of the Association of Computational Linguistics (EACL). 2006.

[167] **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.

[168] 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.

- [169] 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.
- [170] 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

- [171] **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

- [172] **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.
- [173] **Habash, Nizar**. Large Scale Lexeme Based Arabic Morphological Generation. In Proceedings of Traitement Automatique des Langues Naturelles (TALN-04). Fez, Morocco, 2004.
- [174] **Habash, Nizar** and Owen Rambow. Extracting a Tree Adjoining Grammar from the Penn Arabic Treebank. In Proceedings of Traitement Automatique des Langues Naturelles (TALN-04). Fez, Morocco, 2004.
- [175] 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.
- [176] 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

- [177] **Habash, Nizar**. Matador: A Large Scale Spanish-English GHMT System. In Proceedings of the MT Summit, New Orleans, LA, 2003.
- [178] **Habash, Nizar**. System Demonstration Matador: Spanish-English GHMT. In Proceedings of the MT Summit, New Orleans, LA, 2003.
- [179] **Habash, Nizar** and Bonnie Dorr. A Categorical Variation Database for English. In Proceedings of North American Association for Computational Linguistics, Edmonton, Canada, 2003.
- [180] **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

- [181] **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.
- [182] 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.
- [183] **Habash, Nizar**. Generation-Heavy Hybrid Machine Translation. In Proceedings of the International Natural Language Generation Conference (NLG). New York, 2002.

## 2001-1998

- [184] **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.
- [185] **Habash, Nizar**. oxyGen: A Language Independent Linearization Engine. In Proceedings of the Fourth Conference of the Association for Machine Translation in the Americas (AMTA), Cuernavaca, Mexico, 2000.
- [186] **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.
- [187] **Habash, Nizar**. Issues in Palestinian Arabic Spelling Standardization. In the North American Conference on Afro-Asiatic Linguistics (NACAL 27), Baltimore, MD, 1999.
- [188] 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

### 2022

- [189] **Habash, Nizar**, Reham Marzouk, Christian Khairallah, Salam Khalifa. Morphotactic Modeling in an Open-source Multi-dialectal Arabic Morphological Analyzer and Generator. In Proceedings of the SIGMORPHON Workshop on Computational Phonology, Morphology, and Phonetics (@NAACL 2022), Seattle, 2022.
- [190] El-Haj, Mahmoud, Elvis de Souza, Nouran Khallaf, Paul Rayson and **Nizar Habash**. AraSAS: The Open Source Arabic Semantic Tagger. In Proceedings of Workshop on Open-Source Arabic Corpora and Processing Tools (@LREC 2022), Marseille, France, 2022.

### 2021

- [191] Pimentel, Tiago, Maria Ryskina, Sabrina J. Mielke, Shijie Wu, Eleanor Chodroff, Brian Leonard, Garrett Nicolai, Yustinus Ghanggo Ate, Salam Khalifa, **Nizar Habash**, Charbel El-Khaissi, Omer Goldman, Michael Gasser, William Lane, Matt Coler, Arturo Oncevay, Jaime Rafael Montoya Samame, Gema Celeste Silva Villegas, Adam Ek, Jean-Philippe Bernardy, Andrey Shcherbakov, Aziyana Bayyr-ool, Karina Sheifer, Sofya Ganieva, Matvey Plugaryov, Elena Klyachko, Ali Salehi, Andrew Krizhanovsky, Natalia Krizhanovsky, Clara Vania, Sardana Ivanova, Aelita Salchak, Christopher Straughn, Zoey Liu, Jonathan North Washington, Duygu Ataman, Witold Kieraś, Marcin Woliński, Totok Suhardijanto, Niklas Stoehr, Zahroh Nuriah, Shyam Ratan, Francis M. Tyers, Edoardo M. Ponti, Grant Aiton, Richard J. Hatcher, Emily Prud'hommeaux, Ritesh Kumar, Mans Hulden, Botond Barta, Dorina Lakatos, Gábor Szolnok, Judit Ács, Mohit Raj, David Yarowsky, Ryan Cotterell, Ben Ambridge, Ekaterina Vylomova. "SIGMORPHON 2021 Shared Task on Morphological Reinflection: Generalization Across Languages." In Proceedings of the 18th SIGMORPHON Workshop on Computational Research in Phonetics, Phonology, and Morphology, Virtual, 2021.
- [192] Chierici, Alberto, and **Nizar Habash**. A View From The Crowd: Evaluation Challenges for Time-Offset Interaction Applications. In Proceedings of the Workshop on Human Evaluation of NLP Systems (HumEval), Virtual, 2021.
- [193] Abdul-Mageed, Muhammad, Chiyu Zhang, AbdelRahim Elmadany, Houda Bouamor, and **Nizar Habash**. "NADI 2021: The Second Nuanced Arabic Dialect Identification Shared Task." In Proceedings of the Sixth Arabic Natural Language Processing Workshop, 2021.
- [194] Inoue, Go, Bashar Alhafni, Nurpeiis Baimukan, Houda Bouamor, and **Nizar Habash**. "The Interplay of Variant, Size, and Task Type in Arabic Pre-trained Language Models." In Proceedings of the Sixth Arabic Natural Language Processing Workshop, 2021.
- [195] Fadhil, Eryani, and **Habash Nizar**. "Automatic Romanization of Arabic Bibliographic Records." In Proceedings of the Sixth Arabic Natural Language Processing Workshop, 2021.

## 2020

- [196] Taji, Dima and **Nizar Habash**. PALMYRA 2.0: A Configurable Multilingual Platform Independent Tool for Morphology and Syntax Annotation. In Proceedings of the Fourth Workshop on Universal Dependencies (COLING 2020), Online, 2020.
- [197] Shazal, Ali, Aiza Usman, **Nizar Habash**. A Unified Model for Arabizi Detection and Transliteration using Sequence-to-Sequence Models. In Proceedings of the Fifth Arabic Natural Language Processing Workshop (COLING 2020), Online, 2020.
- [198] Abdul-Mageed, Muhammad, Chiyu Zhang, Houda Bouamor, **Nizar Habash**. NADI 2020: The First Nuanced Arabic Dialect Identification Shared Task. In Proceedings of the Fifth Arabic Natural Language Processing Workshop (COLING 2020), Online, 2020.
- [199] Alhafni, Bashar, **Nizar Habash** and Houda Bouamor. Gender-Aware Reinflection using Linguistically Enhanced Neural Models. Proceedings of the Second Workshop on Gender Bias in Natural Language Processing (COLING 2020), Online, 2020.

## 2019

- [200] **Habash, Nizar**, Houda Bouamor and Christine Chung. Automatic Gender Identification and Reinflection in Arabic. In Proceedings of the 1st ACL Workshop on Gender Bias for Natural Language Processing, Florence, Italy, 2019.
- [201] Alshargi, Faisal, Shahd Dibas, Sakhar Alkhereyf, Reem Faraj, Basmah Abdulkareem, Sane Yagi, Ouafaa Kacha, **Nizar Habash** and Owen Rambow. Morphologically Annotated Corpora for Seven Arabic Dialects: Taizi, Sanaani, Najdi, Jordanian, Syrian, Iraqi and Moroccan. In Proceedings of Workshop on Arabic Natural Language Processing, Florence, Italy, 2019.
- [202] Bouamor, Houda, Sabit Hassan, and **Nizar Habash**. The MADAR Shared Task on Arabic Fine-Grained Dialect Identification. In Proceedings of Workshop on Arabic Natural Language Processing, Florence, Italy, 2019.
- [203] Erdmann, Alexander, Salam Khalifa, Mai Oudah, Houda Bouamor, **Nizar Habash**. A Little Linguistics Goes a Long Way: Unsupervised Segmentation with Limited Language Specific Guidance. In Proceedings of SIGMORPHON, Florence, Italy, 2019.

## 2018

- [204] Taji, Dima, Salam Khalifa, Ossama Obeid, Fadhl Eryani and **Nizar Habash**. An Arabic Morphological Analyzer and Generator with Copious Features. In *Proceedings of the SIGMORPHON workshop co-located with the Empirical Methods for Natural Language Processing Conference, Brussels, 2018*.
- [205] Erdmann, Alexander and **Nizar Habash**. Complementary Strategies for Low Resourced Morphological Modeling. In *Proceedings of the SIGMORPHON workshop co-located with the Empirical Methods for Natural Language Processing Conference, Brussels, 2018*.
- [206] Boukaram, Halim-Antoine, **Nizar Habash**, Micheline Ziadee and Majd Sakr. Improving Domain Independent Question Parsing with Synthetic Treebanks. In Proceedings of the COLING Workshop on Linguistic Annotation, Multiword Expressions and Constructions, Santa Fe, USA, 2018.
- [207] Saddiki, Hind, **Nizar Habash**, Violetta Cavalli-Sforza and Muhamed Al-Khalil. Feature Optimization for Predicting Readability of Arabic L1 and L2. In Proceedings of the ACL Workshop on Natural Language Processing Techniques for Educational Applications, Melbourne, Australia, 2018.
- [208] Taji, Dima, Jamila El Gizuli and **Nizar Habash**. An Arabic Dependency Treebank in the Travel Domain. In Proceedings of the Workshop on Free/Open-Source Arabic Corpora and Corpora Processing Tools (OSACT), LREC, Miyazaki, Japan, 2018.
- [209] Badaro, Gilbert, Hussein Jundi, Hazem Hajj, Wassim El-Hajj and **Nizar Habash**. ArSEL: A Large-scale Arabic Sentiment and Emotion Lexicon. In Proceedings of the Workshop on Free/Open-Source Arabic Corpora and Corpora Processing Tools (OSACT), LREC, Miyazaki, Japan, 2018.

## 2017

- [210] Onyibe, Chukwuyem and **Nizar Habash**. OMAM at SemEval-2017 Task 4: English Sentiment Analysis with Conditional Random Fields. International Workshop on Semantic Evaluation (co-located with ACL 2017). Vancouver, 2017.

- [211] 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.
- [212] 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.
- [213] 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.
- [214] 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.
- [215] 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

- [216] 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.
- [217] 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, COLING, Osaka, Japan, 2016.

## 2015

- [218] Saadane, Houda, and **Nizar Habash**. A Conventional Orthography for Algerian Arabic. In Proceedings of the Second Arabic Natural Language Processing Workshop, ACL, Beijing, 2015.
- [219] Hamdi, Ahmed, Alexis Nasr, **Nizar Habash**, and N ria 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.
- [220] Rozovskaya, Alla, Houda Bouamor, **Nizar Habash**, Wajdi Zaghouani, 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.
- [221] 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.
- [222] Zaghouani, 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 Linguistic Annotation Workshop, NAACL, 2015.
- [223] Sajjad, Hassan, Nadir Durrani, Francisco Guzman, Preslav Nakov, Ahmed Abdelali, Stephan Vogel, Wael Salloum, Ahmed El Kholy, **Nizar Habash**. The QCN Egyptian Arabic to English Statistical Machine Translation System for NIST OpenMT'2015. In Proceedings of the Open Machine Translation Evaluation Workshop, Washington DC, 2015.

## 2014

- [224] 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.
- [225] 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.

- [226] 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.
- [227] 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.
- [228] Jebblee, 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.
- [229] Mohit, Behrang, Alla Rozovskaya, **Nizar Habash**, Wajdi Zaghouni and Ossama Obeid. The First QALB Shared Task on Automatic Text Correction for Arabic. In Proceedings of the Arabic Natural Language Processing Workshop, EMNLP, Doha, 2014.
- [230] 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.
- [231] Tomeh, Nadi, **Nizar Habash**, Ramy Eskander and Joseph Le Roux. A Pipeline Approach to Supervised Automatic Error Correction for the QALB-2014 Shared Task. In Proceedings of the Arabic Natural Language Processing Workshop, EMNLP, Doha, 2014.

## 2013

- [232] Seddah, Djamelé, Reut Tsarfaty, Sandra Kübler, Marie Candito, Jinho D. Choi, Richard 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, Yannick Versley, Veronika Vincze, Marcin Woliński, Alina Wroblewska, and Eric Villemonte de la Clergerie. 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.
- [233] 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.
- [234] 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.
- [235] 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

- [236] **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

- [237] 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.
- [238] Marton, Yuval, Ahmed El Kholy 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.

- [239] 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.
- [240] 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

- [241] 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.
- [242] 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.
- [243] 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.
- [244] 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

- [245] 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.
- [246] **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.
- [247] Biadisy, Fadi, Julia Hirschberg and **Nizar Habash**. Spoken Arabic Dialect Identification Using Phonotactic Modeling. In Proceedings of the Workshop on Computational Approaches to Semitic Languages at the meeting of the European Association for Computational Linguistics (EACL), Athens, Greece, 2009.
- [248] 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.

## 2008

- [249] **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.
- [250] 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.
- [251] **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

- [252] Biadsy, 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.
- [253] **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.
- [254] 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.
- [255] 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 Texts. In Proceedings of the Workshop on Beyond Named Entity Recognition: Semantic Labeling for NLP Tasks, LREC, Portugal, 2004.
- [256] 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.
- [257] 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.
- [258] Traum, David and **Nizar Habash**. Generation from Lexical Conceptual Structures. Workshop on Applied Interlinguas, ANLP/NAACL-2000. Seattle, WA.

## Technical Reports

- [1] Noll, Ella, Mai Oudah, and **Nizar Habash**. Simple Automatic Post-editing for Arabic-Japanese Machine Translation. arXiv preprint arXiv:1907.06210.
- [2] Almahairi, Amjad, Kyunghyun Cho, **Nizar Habash**, and Aaron Courville. First result on Arabic neural machine translation. arXiv preprint arXiv:1606.02680. 2016.
- [3] **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.
- [4] 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.
- [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] 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.

- [10] 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.
- [11] **Habash, Nizar**. Generation-Heavy Hybrid Machine Translation. Doctoral Dissertation. Computer Science Department, University of Maryland College Park, 2003.
- [12] **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.
- [13] **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] **Habash, Nizar**. [An Introduction to Semitish](#). Self-published book manuscript, 2015.
- [2] **Habash, Nizar**. "On Arabic and its Dialects," Multilingual Magazine. #81 Volume 17 Issue 5, 2006.
- [3] **Habash, Nizar**. [Introduction to Delason: The Complete Guide to the Artificial Language](#). Self-published book manuscript, First edition 1998. Second edition 2012.

## Oral Presentations

---

### Invited Talks

- [1] **Habash, Nizar**. Arabic and Technology. Invited discussion as part of the Arabic Language Summit. Dubai, Dec 19, 2021.
- [2] **Habash, Nizar**. Arabic Natural Language Processing at CAMEL Lab. Invited Talk, Al Ain University, Sep 23, 2021.
- [3] **Habash, Nizar**. A Thousand and One Sources of Language Processing Bias. Invited Keynote Speech. Workshop on Gender Bias in Natural Language Processing, Aug 5, 2021.
- [4] **Habash, Nizar**. Computational Modeling of Arabic Dialects. Invited Talk. Amazon, Jun 5, 2021.
- [5] **Habash, Nizar**. Arabic and AI: Challenges and Opportunities. Invited Talk. MBZUAI, May 25, 2021.
- [6] **Habash, Nizar**. Arabic Natural Language Processing at CAMEL Lab. Invited Talk. PSUT, May 24, 2021.
- [7] **Habash, Nizar**. A Short Introduction to Arabic Natural Language Processing. Online, KAUST Webinar on Machine Learning and Arabic NLP, Nov 30, 2020.
- [8] **Habash, Nizar**, and Obeid, Ossama. CAMEL Tools: an open-source Python Arabic Natural Language Processing (NLP) toolkit. (Original title: مجموعة أدوات كامل: مجموعة أدوات مفتوحة المصدر بلغة (بايثون لمعالجة اللغة العربية). IWAN Research Group. Nov 11, 2020.
- [9] **Habash, Nizar**. Automatic Gender Identification and Reinflection in Arabic. Invited Talk, QCRI/ALT, Sep 29, 2020.
- [10] **Habash, Nizar**. Artificial Intelligence and the Arabic Language: Challenges and Solutions. (Original title: "النكاء الاصطناعي واللغة العربية: تحديات وحلول") Harvard Business Review Arabia Webinar. June 23, 2020.
- [11] **Habash, Nizar**. Arabic NLP Insights for OCR and Historical Document Analysis. University of Maryland College Park OCR and Digital Text Production Workshop. Invited Talk. January 30, 2020.
- [12] **Habash, Nizar**. Fine-Grained Arabic Dialect Identification. University of Maryland College Park. Invited Talk. January 28, 2020.
- [13] **Habash, Nizar**. Arabic Natural Language Processing: Challenges and Solutions. Grammarly, Inc. Kiev, Ukraine, March 26, 2019.
- [14] **Habash, Nizar**. On Arabic and Artificial Intelligence. International Symposium on Robotics and Artificial Intelligence, New York University Abu Dhabi, Abu Dhabi, March 11, 2019.
- [15] **Habash, Nizar**. Computational Modeling of Arabic and its Dialects. New York University Abu Dhabi, Identity Politics Group Meeting, Abu Dhabi, November 22, 2018.

- [16] **Habash, Nizar.** Multi-Arabic Dialect Applications and Resources. Paris 13 University, Paris, July 2, 2018.
- [17] **Habash, Nizar.** Research Efforts at the Computational Approaches to Modeling Language Lab. Paris 13 University, Paris, June 27, 2018.
- [18] **Habash, Nizar.** Morphological Annotation of Gulf Arabic. Invited Speaker. UAEU Linguistics Seminar. Al-Ain, April 10, 2018.
- [19] **Habash, Nizar.** Language Technologies for Arabic and its Dialects. Keynote Speaker. Arab Women in Computing Conference. Beirut. 10 Aug 2017.
- [20] **Habash, Nizar.** Computational Processing of Arabic for Digital Humanistic Research. Digital Humanities Abu Dhabi DHAD conference. Abu Dhabi. 12 April 2017.
- [21] **Habash, Nizar.** Language Technologies for Arabic and its Dialects. NYUAD Institute Public talk. November 15, 2016. Abu Dhabi.
- [22] **Habash, Nizar.** Computational Processing of the Arabic Language. Invited Talk. Nazarbayev University, Astana, Kazakhstan, October 20, 2016.
- [23] **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.
- [24] **Habash, Nizar.** Morphological Processing of Arabic and its Dialects. Invited Talk. University of Malta, Malta, April 22, 2016.
- [25] **Habash, Nizar.** Computational Processing of Gulf Arabic. Invited Talk. British University of Dubai, Dubai, UAE, April 13, 2016.
- [26] **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).
- [27] **Habash, Nizar.** Computational Processing of Arabic Dialects. Invited Talk. Université Toulouse III - Paul Sabatier., Toulouse, France, May 11, 2015.
- [28] **Habash, Nizar.** Arabic and Arabic Dialect Technologies. Invited Speaker. NYU Abu Dhabi Tech Fest, April 25, 2015.
- [29] **Habash, Nizar.** Arabic Language Technology: Current Status and Future Potential. Invited Speaker Defining Standards Conference, Abu Dhabi, UAE, November 19, 2014.
- [30] **Habash, Nizar.** Morphological Processing of Arabic and its Dialects. Computer Science Seminar Series. NYU Abu Dhabi, November 13, 2014.
- [31] **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.
- [32] **Habash, Nizar.** Morphological Analysis and Disambiguation for Dialectal Arabic. Invited talk. IBM Professional Interest Community Seminar, June 26, 2013.
- [33] **Habash, Nizar.** Arabic Natural Language Processing: Challenges, Solutions, Applications. Invited talk. New York University Abu Dhabi, May 2, 2013.
- [34] **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.
- [35] **Habash, Nizar.** Arabic Natural Language Processing: Challenges, Solutions, Applications. Invited talk. Invited talk. Polytechnic Institute of New York University, April 11, 2013.
- [36] **Habash, Nizar.** Arabic Natural Language Processing: Challenges, Solutions, Applications. Invited talk. Invited talk. Courant Institute, New York University, April 10, 2013.
- [37] **Habash, Nizar.** Arabic Natural Language Processing: Challenges, Solutions, Applications. Invited talk. Invited talk. Language Technologies Institute. Carnegie Mellon University, April 4, 2013.
- [38] **Habash, Nizar.** Arabic Natural Language Processing: Challenges, Solutions, Applications. Invited talk. Invited talk. Computer and Information Technology Department, Purdue University, April 2, 2013.
- [39] **Habash, Nizar.** Arabic Dialect Processing and Speech-to-Speech Machine Translation. Workshop on Simultaneous Interpretation and Multimodal Translation. Invited talk. ATT, March 25, 2013.

- [40] **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.
- [41] **Habash, Nizar.** Arabic Parsing and Statistical Machine Translation. Invited Talk, Language Technology Institute. Carnegie Mellon University, July 2012.
- [42] **Habash, Nizar.** Arabic Computational Linguistics: Past, Present and Future. Keynote Presentation. Arabic Linguistics Symposium. Columbia University, March 2012.
- [43] **Habash, Nizar.** Arabic Morphological Analysis and Disambiguation: Problem, Solution, Applications. Invited Talk. Linguistics Speaker Series. Georgetown University, Washington D.C., 2009.
- [44] **Habash, Nizar.** Automatic Diacritization of Arabic Text. Invited Talk. Columbia Linguistic Society, Columbia University, New York, 2009.
- [45] **Habash, Nizar.** Arabic Diacritization through Full Morphological Tagging. Invited Talk, Language Technology Institute Seminar Series. Carnegie Mellon University, 2007.
- [46] **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.
- [47] **Habash, Nizar.** Disambiguation of Rich Arabic Morphological Analyses. Invited Talk, NYU Natural Language Processing Colloquium. April 14, 2006.
- [48] **Mona Diab, Nizar Habash,** and Owen Rambow. Arabic Dialect Parsing. Invited Talk, Linguistic Data Consortium, University of Pennsylvania, December 9, 2005
- [49] **Habash, Nizar.** Combining Symbolic and Statistical Techniques for Machine Translation and Morphological Disambiguation. Invited Talk, AT&T Labs, Florham Park, NJ, October 7, 2005.
- [50] **Habash, Nizar.** Generation-Heavy Hybrid Machine Translation. Invited Talk. IBM TJ Watson Research Center, March 9, 2005.
- [51] **Habash, Nizar.** Generation-Heavy Hybrid Machine Translation. Invited Talk. Computer Science Department Seminar, City College of New York, March 8, 2005.
- [52] **Habash, Nizar.** Generation-Heavy Hybrid Machine Translation. Invited Talk. Language Technologies Institute Seminar. Carnegie Mellon University, November 19, 2004.
- [53] **Habash, Nizar.** Semitic Linguistic Phenomena. Invited Talk. Workshop on Machine Translation for Semitic Languages, MT Summit, New Orleans, LA, 2003.

### Invited Panels

- [1] **Habash, Nizar.** Computational Tools for Research in the Arabic Script Humanities. Digital Humanities and Islamic Studies Online Workshop. Invited Panelist. Oxford Center for Islamic Studies. Nov 20, 2021.
- [2] **Habash, Nizar.** Gender and Bias in a Multicultural and Multilingual Context. Invited Panelist. Workshop on Gender Bias in Natural Language Processing, Aug 5, 2021.
- [3] **Habash, Nizar.** Experiences, Challenges, and Opportunities for Collaboration in Industry and Academia. Invited Panelist. Aixplain. May 26, 2021.
- [4] **Habash, Nizar.** اللغة العربية: أين مكانها في عالم الرقمنة؟ Invited Panelist. Ministry of Culture and Knowledge Development. Dec 22, 2020.
- [5] **Habash, Nizar.** Revisiting Nuun: A solution to Arabic display from 20 years ago. Invited Panelist. Right-to-Left Unconference. Winter Institute in Digital Humanities (WIDH 2020), New York University Abu Dhabi, January 21, 2020.
- [6] **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
- [7] **Habash, Nizar.** Treebanking for Machine Translation. Invited Panel Talk, GALE PI Meeting, Tampa, Florida, 2008.
- [8] **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.

- [9] **Habash, Nizar.** Hybrid Machine Translation. Invited Panel Talk. Conference of the Association for Machine Translation in the Americas, AMTA, 2006.

### Conference Presentations, Colloquia, Forums, Symposia, Workshops, Bake-offs

- [1] Palfreyman, David M. and **Nizar Habash.** An Arabic-English bilingual writer corpus for research on biliteracy. Experimental Arabic Linguistics Conference (EXAL), Virtual, 2021.
- [2] Inoue, Go, Akihiko Kato, **Nizar Habash,** Yuji Matsumoto. A Cross-Lingual Visualization Tool for Real-Time Tweets. The 13th Symposium for Young Researcher Association for NLP Studies. Kagawa, Japan, 2018.
- [3] 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 Poster.
- [4] Zaghouani, 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. Invited Poster.
- [5] 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.
- [6] Ahmed El Kholy and Nizar Habash. English Language Bridging for Persian-Arabic and Hebrew-Arabic Statistical Machine Translation. NYUAD Annual Research Conference, 2015.
- [7] 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.
- [8] 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.
- [9] 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.
- [10] 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.
- [11] 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.
- [12] 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.
- [13] 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.
- [14] **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.
- [15] **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.
- [16] Marton, Yuval, **Nizar Habash,** Owen Rambow and Sarah Alkuhlani. SPMRL'13 Shared Task System: The CADIM Arabic Dependency Parser. 2013.
- [17] El Kholy, Ahmed, Wael Salloum and **Nizar Habash.** Columbia University 2012 Arabic-English, Farsi-English and Dari-English MT Evaluation Systems. NIST Machine Translation Evaluation, 2012.
- [18] Carpuat, Marine, Mahmoud Ghoneim, **Nizar Habash** and Mona Diab. Columbia University 2009 Arabic-English MT Evaluation System. NIST Machine Translation Evaluation, 2009.

- [19] Diab, Mona, **Nizar Habash** and Owen Rambow. Tools for Arabic Morphological Analysis, Disambiguation and Tokenization. Foreign Language Technology Forum, July 2008.
- [20] **Habash, Nizar**. Techniques for Online OOV Handling in Phrase-based Statistical Machine Translation. *Columbia University NLP Colloquium*, April 2008.
- [21] **Habash, Nizar**. Five Techniques for OOV Reduction in Arabic-English and Urdu-English SMT. *NIST MTEval Workshop*, March 2008.
- [22] **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.
- [23] **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.
- [24] 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.
- [25] **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.
- [26] 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.

## Resources

---

### Tools

- [1] **Arabic Romanization ALA-LC**: Automatic Romanization of Arabic Bibliographic Records in the style of the Library of Congress (2021, with Fadhil Eryani)
  - ◇ Paper: [PDF](#)
  - ◇ Demo: <http://romanize-arabic.camel-lab.com/>
  - ◇ Github: [https://github.com/CAMeL-Lab/Arabic\\_ALA-LC\\_Romanization](https://github.com/CAMeL-Lab/Arabic_ALA-LC_Romanization)
- [2] **CAMEL Tools**: An Open Source Python Toolkit for Arabic Natural Language Processing. (2020, with Ossama Obeid and many Camel Lab members)
  - ◇ Paper: [PDF](#)
  - ◇ Github: [https://github.com/CAMeL-Lab/camel\\_tools](https://github.com/CAMeL-Lab/camel_tools)
- [3] **The Readability Leveled Arabic Thesaurus**: A tool for smart search over related words in Arabic (2020, developed with Muhamed Al Khalil and Ross Jiang).
  - ◇ Demo: <http://samer.camel-lab.com/>
- [4] **CALIMA Star**: A rich Arabic morphological analyzer and generator (2019, developed with Dima Taji, Salam Khalifa and Ossama Obeid)
  - ◇ Paper: [PDF](#)
  - ◇ Demo: <https://calimastar.abudhabi.nyu.edu/analyzer/>
- [5] **Palmyra**: A Platform Independent Dependency Annotation Tool for Morphologically Rich Languages (2018, developed with Talha Javed and Dima Taji)
  - ◇ Paper: [PDF](#)
  - ◇ Demo: <http://palmyra.camel-lab.com/>
- [6] **ADIDA**: Automatic Dialect Identification for Arabic. (2018-2019, developed with Ossama Obeid, Muhammad Salameh, and Houda Bouamor.)
  - ◇ Paper: [PDF](#), [PDF](#)
  - ◇ Demo: <https://adida.abudhabi.nyu.edu/#/>
  - ◇ Contact [nizar.habash@nyu.edu](mailto:nizar.habash@nyu.edu) if interested in getting access to the tool.
- [7] **MADARI**: A Web Interface for Joint Arabic Morphological Annotation and Spelling Correction. (2018, developed with Ossama, Obeid, Salam Khalifa, Houda Bouamor, Wajdi Zaghouni and Kemal Oflazer.)
  - ◇ Paper: [PDF](#)

- ◇ Contact nizar.habash@nyu.edu if interested in getting access to the tool.
- [8] **QUTR**: Query Translator -- A Cross-lingual Messenger with Keyword Searchable Phrases for the Travel Domain. (2018, developed with Shehroze Khan, Jihyun Kim, Tarik Zulfikarpasic, and Peter Chen.)
  - ◇ Paper: [PDF](#)
  - ◇ Demo: [Video](#)
  - ◇ Contact nizar.habash@nyu.edu if interested in getting access to the tool.
- [9] **TOIA**: Time-offset Interaction Application -- A Bilingual Interactive Human Avatar Dialogue System. (2018, developed with Dana Abu Ali, Muaz Ahmad, Hayat Al Hassan, Paula Dozsa, Ming Hu, and Jose Varias.)
  - ◇ Paper: [PDF](#)
  - ◇ Demo: [Video \(Arabic-English\)](#), [Video \(English-English\)](#);
  - ◇ Contact nizar.habash@nyu.edu if interested in getting access to the tool.
- [10] **Gulf Arabic Verb Analyzer** (2017, developed with Salam Khalifa and Sara Hassan)
  - ◇ Paper: [PDF](#)
  - ◇ Download: <https://camel.abudhabi.nyu.edu/resources/>
- [11] **CamelParser**: A System for Arabic Syntactic Analysis and Morphological Disambiguation (2016, developed with Anas Shahrour, Salam Khalifa and Dima Taji)
  - ◇ Paper: [PDF](#)
  - ◇ Download: <https://camel.abudhabi.nyu.edu/resources/>
  - ◇ *CamelParser is the latest most improved version of Arabic Dependency parsing in the Columbia Arabic Treebank style. Previous versions are not supported any more: Columbia Arabic Treebank Dependency Parser (2013), in collaboration with Yuval Marton, Owen Rambow, Ryan Roth and Sarah Alkuhlani*
- [12] **Botta**: An Egyptian Arabic Dialect Chatbot (2016, developed with Dana Abu Ali)
  - ◇ Paper: [PDF](#)
  - ◇ Download: <https://camel.abudhabi.nyu.edu/resources/>
- [13] **YAMAMA**: Yet Another Multi-Dialect Arabic Morphological Analyzer (2016, developed with Salam Khalifa and Nasser Zalmout)
  - ◇ Paper: [PDF](#)
  - ◇ Download: <https://camel.abudhabi.nyu.edu/resources/>
- [14] **DALILA**: The Dialectal Arabic Linguistic Learning Assistant -- A Chrome Extension. (2016, developed with Salam Khalifa and Houda Bouamor)
  - ◇ Paper: [PDF](#)
  - ◇ Download: <https://camel.abudhabi.nyu.edu/resources/>
- [15] **ADAM**: Analyzer for dialectal Arabic Morphology (2014, developed with Wael Salloum)
  - ◇ Paper: [PDF](#)
  - ◇ Download: <http://www.nizarhabash.com/Columbia/Tools.html>
- [16] **MADAMIRA**: A system for morphological analysis and disambiguation, base phrase chunking and named entity recognition. (2014, as part of the Columbia NADIA team; MADAMIRA merges the popular MADA tool [Habash and Rambow] and AMIRA tool [Diab])
  - ◇ Paper: [PDF](#)
  - ◇ Demo: <https://camel.abudhabi.nyu.edu/madamira/> (English interface)
  - ◇ Demo: <https://camel.abudhabi.nyu.edu/madamira/?locale=ar> (Arabic interface)
  - ◇ Download: ([click here](#))
  - ◇ *MADAMIRA is the latest version in a series of tools that are no longer supported:*
    - i. *MADA+TOKAN: Morphological Analysis and Disambiguation for Arabic (2005, developed with Owen Rambow and Ryan Roth)*
    - ii. *MADA-ARZ 0.3: Morphological Analysis and Disambiguation for Arabic (Egyptian version) (2012, developed with Owen Rambow and Ryan Roth)*
- [17] **3ARRIB**: A system for transliteration from Romanized Arabic to Arabic script (2014, developed with Mohamed Al-Badrashiny, Ramy Eskander, and Owen Rambow)
  - ◇ Paper: [PDF](#), [PDF](#)
  - ◇ Demo: <https://camel.abudhabi.nyu.edu/arrrib/>
- [18] **DIRA**: Dialectal Arabic Information Retrieval Assistant (2013, as part of the Columbia NADIA team; DIRA effort led by Nizar Habash and Owen Rambow)
  - ◇ Paper: [PDF](#)

- ◇ Demo: <https://camel.abudhabi.nyu.edu/dira/>
- [19] **CODAFY**: Conventional Orthography of Dialectal Arabic Normalizer (2013, developed with Ramy Eskander and Owen Rambow)
  - ◇ Paper: [PDF](#)
  - ◇ Demo: <http://www.nizarhabash.com/Columbia/Tools.html>
- [20] **MAGEAD**: Morphological Analyzer/Generator for Arabic and its Dialects (2006-2013, developed with Owen Rambow, Mohammed Altantawy)
  - ◇ Paper: [PDF](#), [PDF](#)
  - ◇ Download: <http://www.nizarhabash.com/Columbia/Tools.html>
- [21] **ELISSA**: Dialectal Arabic to Modern Standard Arabic Translation System (2012, developed with Wael Salloum)
  - ◇ Paper: [PDF](#), [PDF](#)
  - ◇ Download: <http://www.nizarhabash.com/Columbia/Tools.html>
- [22] **Arabic C2D**: Arabic Constituency-to-Dependency Conversion tool (2009)
  - ◇ Paper: [PDF](#)
  - ◇ Download: <http://www.nizarhabash.com/Columbia/Tools.html>
- [23] **SEPIA**: Machine Translation Evaluation Metric (2008, developed with Ahmed El Kholy)
  - ◇ Paper: [PDF](#)
  - ◇ Download: <http://www.nizarhabash.com/Columbia/Tools.html>
- [24] **REMOOV**: Tool for Reducing Out-Of-Vocabulary in Machine Translation (2007)
  - ◇ Paper: [PDF](#), [PDF](#)
  - ◇ Download: <http://www.nizarhabash.com/Columbia/Tools.html>

## Corpora and Lexicons

- [1] **MADAR CODA Corpus**: A Spelling Correction Corpus for Multiple Arabic Dialects (2020, with Fadhl Eryani).
  - ◇ Paper: [PDF](#)
  - ◇ Download: <https://camel.abudhabi.nyu.edu/madar-coda-corpus/>
- [2] **SAMER Readability Lexicon**: A 26,000 lemma resource labelled for readability level (2020, with Muhamed Al Khalil)
  - ◇ Paper: [PDF](#)
  - ◇ Download: <https://camel.abudhabi.nyu.edu/samer-readability-lexicon/>
  - ◇ Demo: <http://samer.camel-lab.com/>
- [3] **Margarita Dialogue Corpus**: a corpus of dialogues paired with a database of videos (2020, with Alberto Chierici, Alberto and Margarita Bicec)
  - ◇ Paper: [PDF](#)
  - ◇ Download: <https://camel.abudhabi.nyu.edu/margarita-dialogue-corpus/>
- [4] **The Arabic Parallel Gender Corpus**: A corpus annotated for speaker grammatical gender (Masc/Fem/Both) and paired with a reinflected version in the other gender (Fem/Masc). (2019, with Houda Bouamor and Christine Chung)
  - ◇ Paper: [PDF](#)
  - ◇ Download: <https://camel.abudhabi.nyu.edu/arabic-parallel-gender-corpus/>
- [5] **MADAR Dialect ID Corpus**: A data set for dialect identification adding Twitter user labels on top of the MADAR corpora.
  - ◇ Paper: <https://www.aclweb.org/anthology/W19-4622.pdf>
  - ◇ Download: <https://camel.abudhabi.nyu.edu/madar-shared-task-2019/>
- [6] **Arabic Multidialectal Word Embeddings**: Pre-trained word embeddings learned from different dialects of Arabic. (2018, developed with Alexander Erdmann and Nasser Zalmout)
  - ◇ Paper: [PDF](#)
  - ◇ Download: <https://camel.abudhabi.nyu.edu/resources/>
- [7] **MADAR Corpus**: A parallel corpus of 25 city Arab dialects in the travel domain. (2018, developed with Houda Bouamor, Mohammad Salameh, and many others)
  - ◇ Paper: [PDF](#)
  - ◇ Website: <https://camel.abudhabi.nyu.edu/madar/>

- [8] **MADAR Lexicon**: A dictionary covering 25 city Arab dialects. (2018, developed with Houda Bouamor, Mohammad Salameh, and many others)
- ◇ Paper: [PDF](#)
  - ◇ Website: <https://camel.abudhabi.nyu.edu/madar/>
- [9] **SAMER Readability Corpus**: A Leveled Reading Corpus of Modern Standard Arabic. (2018, developed with Muhamed Al Khalil, Hind Saddiki and Latifa Alfalasi)
- ◇ Paper: [PDF](#)
  - ◇ Website: <https://camel.abudhabi.nyu.edu/samer-readability-lexicon/>
- [10] **Annotated Gumar Corpus**: A Morphologically Annotated Corpus of Emirati Arabic. (2018, developed with Salam Khalifa, Fadhl Eryani, Ossama Obeid, Dana Abdulrahim and Meera Al Kaabi)
- ◇ Paper: [PDF](#)
  - ◇ Download: <https://camel.abudhabi.nyu.edu/resources/>
- [11] **PalmTreebank – Part1 Travel Domain**: An Arabic Dependency Treebank in the Travel Domain. (2018, developed with Dima Taji and Jamila El Gizuli)
- ◇ Paper: [PDF](#)
  - ◇ Download: <https://camel.abudhabi.nyu.edu/resources/>
- [12] **ArSEL**: A Large-scale Arabic Sentiment and Emotion Lexicon. (2018, developed with Gilbert Badaro, Hussein Jundi, Hazem Hajj and Wassim El-Hajj)
- ◇ Paper: [PDF](#)
  - ◇ Download: [http://oma-project.azurewebsites.net/ArSenL/ArSEL\\_intro](http://oma-project.azurewebsites.net/ArSenL/ArSEL_intro)
- [13] **Japanese-Arabic News Corpus**: A Parallel Corpus of Arabic-Japanese News Articles. (2018, developed with Go Inoue, Yuji Matsumoto and Hiroyuki Aoyama)
- ◇ Paper: [PDF](#); Contact the paper authors if interested in getting access to the data.
- [14] **Arabic Question Corpus**: A corpus of manually and synthetically constructed syntactic analyses of questions in Arabic. (2018, developed with Halim-Antoine Boukaram, Micheline Ziadee, and Majd Sakr)
- ◇ Paper: [PDF](#)
  - ◇ Contact the authors to get access to the data.
- [15] **Arab-Acquis**: Parallel Corpus of Arabic and European Languages (2017, developed with Dima Taji and Nasser Zalmout)
- ◇ Paper: [PDF](#)
  - ◇ Download: <https://camel.abudhabi.nyu.edu/resources/>
- [16] **NUDAR**: NYU Universal Dependencies for Arabic. (2017, developed with Dima Taji)
- ◇ Paper: [PDF](#)
  - ◇ Download: <https://camel.abudhabi.nyu.edu/resources/>
- [17] **ArabScribe**: A dataset of 10,000 transcriptions of written impressions of what Arabic words sound like. (2017, developed with Lingliang Zhang)
- ◇ Paper: [PDF](#)
  - ◇ Download: <https://camel.abudhabi.nyu.edu/resources/>
- [18] **Gumar Corpus**: The Gulf Arabic Corpus – A 100 Million word dataset (2016, developed with Salam Khalifa)
- ◇ Paper: [PDF](#)
  - ◇ Website: <https://camel.abudhabi.nyu.edu/gumar/>
  - ◇ Download language models and n-grams: <https://camel.abudhabi.nyu.edu/resources/>
- [19] **MAMAC**: The Moroccan Arabic Morphologically Annotated Corpus. (2016, developed with Aidan Kaplan, Faisal Al-Shargi, Ramy Eskander, and Owen Rambow)
- ◇ Paper: [PDF](#)
  - ◇ Contact [Owen Rambow](#) to get access to the data
- [20] **SYAMAC**: The Sanaani Yemeni Arabic Morphologically Annotated Corpus. (2016, developed with Faisal Al-Shargi, Ramy Eskander, and Owen Rambow)
- ◇ Paper: [PDF](#)
  - ◇ Contact [Owen Rambow](#) to get access to the data
- [21] **SIMMR**: Simplified Ingredient Merging Map in Recipes -- A data set and algorithm for parsing recipes (2015, developed with Jermsak Jermsurawong)
- ◇ Paper: [PDF](#)

- ◇ Download: <https://camel.abudhabi.nyu.edu/resources/>
- [22] **Curras**: A Morphologically Annotated Corpus of Palestinian Arabic. (2015, developed with Mustafa Jarrar (PI), Faeq Alrimawi, Diyam Akra and Nasser Zalmout)
  - ◇ Paper: [PDF](#)
  - ◇ Download: <http://portal.sina.birzeit.edu/curras/>
- [23] **arTenTen**: Corpus of the Arabic Web. (2014, developed with Adam Kilgarriff (PI), Tressy Arts, Yonatan Belinkov and Vit Suchomel)
  - ◇ Paper: [PDF](#)
  - ◇ Website: <https://www.sketchengine.eu/artenten-arabic-corpus/>
- [24] **QALB**: The Qatar Arabic Language Bank. (2014-2015, developed with Behrang Mohit, Wajdi Zaghrouani, Kemal Oflazer and many others)
  - ◇ Papers: [PDF](#), [PDF](#), [PDF](#)
  - ◇ Website: <http://nlp.qatar.cmu.edu/qalb/>
  - ◇ Download: Please Register [here](#); and email [nizar.habash@nyu.edu](mailto:nizar.habash@nyu.edu)
- [25] **ArSenL**: A Large-scale Standard Arabic Sentiment Lexicon. (2014, developed with Gilbert Badaro, Ramy Baly, Hazem Hajj and Wassim El-Hajj)
  - ◇ Paper: [PDF](#)
  - ◇ Download: [http://oma-project.azurewebsites.net/ArSenL/download\\_intro](http://oma-project.azurewebsites.net/ArSenL/download_intro)
- [26] **MultiDial Corpus**: A Multidialectal Parallel Corpus of Arabic. (2014, developed with Houda Bouamor and Kemal Oflazer)
  - ◇ Paper: [PDF](#)
  - ◇ Contact: [Houda Bouamor](#) to get access to the data.
- [27] **Tharwa**: A Large Scale Dialectal Arabic - Standard Arabic - English Lexicon. (2014, developed with Mona Diab (lead) and many others)
  - ◇ Paper: [PDF](#)
  - ◇ Contact [Mona Diab](#) if interested in getting access to the data.
- [28] **Arabizi-Arabic Corpus**: The BOLT Egyptian Arabic SMS/Chat and Transliteration. (2014, developed with Mohamed Maamouri, Ann Bies, Zhiyi Song, and many others)
  - ◇ Paper: [PDF](#)
  - ◇ Download: <https://catalog.ldc.upenn.edu/LDC2017T07>
- [29] **ArabicGNR Dataset**: Database for Computational Modeling of Arabic Nominal Gender, Number and Rationality. (2013, developed with Sarah Alkuhlani)
  - ◇ Paper: [PDF](#)
  - ◇ Download: [Columbia Technology Ventures](#)
- [30] **CATiB 1.0**: The Columbia Arabic Treebank. (2009)
  - ◇ Paper: [PDF](#)
  - ◇ Contact the [LDC](#) to get the data.
- [31] **CATVAR**: The Categorical Variation Database. (2003, developed with Bonnie Dorr)
  - ◇ Paper: [PDF](#)
  - ◇ Website: <https://clipdemos.umiacs.umd.edu/catvar/>
  - ◇ Contact: [nizar.habash@nyu.edu](mailto:nizar.habash@nyu.edu)

## Teaching Experience

---

### Classes Taught

#### *New York University Abu Dhabi*

- [1] Undergraduate Computer Science Course. CS-UH 3090: Research Seminar in Computer Science, Spring 2021. 39 students. [\[Syllabus\]](#)
- [2] Undergraduate Computer Science Course. CS-UH 3090: Research Seminar in Computer Science, Spring 2020. 46 students. [\[Syllabus\]](#)
- [3] Undergraduate Core Course. CADT-UH 1011: Words. New York University Abu Dhabi, Fall 2019. 16 students. [\[Syllabus\]](#)

- [4] Undergraduate Computer Science Course. CS-UH 2216: Natural Language Processing. New York University Abu Dhabi, Spring 2019. 23 students. [[Syllabus](#)]
- [5] Undergraduate Core Course. CADT-UH 1011: Words. New York University Abu Dhabi, Fall 2018. 15 students. [[Syllabus](#)]
- [6] Undergraduate Core Course. CADT-UH 1011: Words. New York University Abu Dhabi, Fall 2017. 15 students. [[Syllabus](#)]
- [7] Undergraduate Computer Science Course. CS-AD 220: Natural Language Processing. New York University Abu Dhabi, Spring 2017. 11 students. [[Syllabus](#)]
- [8] Undergraduate Core Course. COREI-AD 58: Words. New York University Abu Dhabi, Fall 2016. 14 students. [[Syllabus](#)]
- [9] Undergraduate Computer Science Course. CS-AD 220: Natural Language Processing. New York University Abu Dhabi, Spring 2016. 12 students. [[Syllabus](#)]
- [10] Undergraduate Core Course. COREI-AD 58: Words. New York University Abu Dhabi, Fall 2015. 14 students. [[Syllabus](#)]
- [11] Undergraduate Computer Science Course. CS-AD 219: Topics in Computer Science: Natural Language Processing. New York University Abu Dhabi, Spring 2015. 3 students. [[Syllabus](#)]

### ***Columbia University***

- [12] 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/>]
- [13] Instructor of graduate-level course. E6998 section 003: Machine Translation. Department of Computer Science, Columbia University, Fall 2011. 9 students.
- [14] Instructor of graduate-level course. E6998 section 003: Machine Translation. Department of Computer Science, Columbia University, Fall 2010. 21 students.

### ***University of Maryland College Park***

- [15] 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] Sébastien Jean, New York University, 2021 (defense)
- [2] Alexander Erdmann, The Ohio State University, 2018 (proposal), 2020 (defense)
- [3] Gilbert Badaro, American University of Beirut, 2016 (proposal), 2019 (defense)
- [4] Nasser Zalmout, New York University, 2018 (proposal), 2019 (defense) [my advisee]
- [5] Sawsan Alqahtani, The George Washington University, 2019 (defense)
- [6] Alberto Chierici, New York University Abu Dhabi, 2019 (qualifier) [my advisee]
- [7] Alexander Erdmann, Ohio State University, 2018 (candidacy)
- [8] Wael Salloum, Columbia University, 2014 (candidacy), 2018 (defense) [my advisee]
- [9] Hamzah Luqman, King Fahd University of Petroleum & Minerals, 2017 (defense)
- [10] Ramy Baly, American University of Beirut, 2014 (proposal), 2016 (defense)
- [11] Leena Lulu, United Arab Emirates University, 2016 (defense)
- [12] Nawar Halabi, University of South Hampton, 2016 (defense)
- [13] Claudia Borg, University of Malta, 2016 (defense)
- [14] Mohammad Salameh, University of Alberta, 2016 (defense)

- [15] Ahmed El Kholy, Columbia University, 2013 (proposal), 2015 (defense) [my advisee]
- [16] Walid Aransa, L'Université du Maine, 2015 (defense)
- [17] Iskandar Keskes, L'Institut de Recherche en Informatique de Toulouse, 2015 (defense)
- [18] Weiyun Ma, Columbia University, 2012 (candidacy), 2013 (proposal), 2014 (defense)
- [19] Kais Dukes, Leeds University, 2013 (defense) [secondary advisor]
- [20] Sarah Alkuhlani, Columbia University, 2013 (candidacy) [my advisee]
- [21] Kristen Parton, Columbia University, 2009 (candidacy), 2010 (proposal), 2012 (defense).
- [22] Sara Stymne, Linköping University, 2012 (defense; served as opponent)
- [23] Fadi Biadsy, Columbia University, 2008 (candidacy), 2010 (proposal), 2011 (defense).
- [24] Jakob Elming, Copenhagen Business School, 2009 (defense).
- [25] Josep Crego, Universitat Politècnica de Catalunya, 2008 (defense).
- [26] Otakar Smrz, Charles University, Czech Republic, 2007 (defense).
- [27] Andrew Rosenberg, Columbia University, 2006 (candidacy).

## **Masters Committees**

- [1] Obeida Eljundi, American University of Beirut, 2018 (proposal), 2021 (defense)
- [2] Alaa Maarouf, American University of Beirut, 2018 (proposal), 2021 (defense)
- [3] Diyam Akra, Birzeit University, Computer Science, 2015 (defense)
- [4] Tejaswini Pepadati, Columbia University, Computer Science, 2014 (defense; external)
- [5] Ibrahim Saleh, Georgetown University, Linguistics, 2009 (Reader)

## **Undergraduate Academic Mentoring**

- **New York University Abu Dhabi**

- [1] Bhavicka Mohta (Class of 2024)
- [2] Dariga Shokayeva (Class of 2024)
- [3] Dixit Timilsina (Class of 2024)
- [4] Sagynbek Talgatuly (Class of 2024)
- [5] Shyngys Karishev (Class of 2024)
- [6] Soumen Mohanty (Class of 2024)
- [7] Ahmed Elshabrawy (Class of 2023)
- [8] Brian Shin (Class of 2023)
- [9] Gopika Krishnan (Class of 2023)
- [10] Hyun Woo Lim (Class of 2023)
- [11] Mariam Elgamal (Class of 2023)
- [12] Ons Taktak (Class of 2023)
- [13] Shaurya Singh (Class of 2023)
- [14] Enid Mollel (Class of 2022)
- [15] Lachlan Pham (Class of 2022)
- [16] Nadja Fejzic (Class of 2022)
- [17] Nhi Pham Le Yen (Class of 2022)
- [18] Nurpeiis Baimukan (Class of 2022)
- [19] Runyao Fan (Class of 2022)
- [20] Sangjin Lee (Class of 2022)
- [21] Usman Ali (Class of 2022)
- [22] Arb Mehmeti (Class of 2021)
- [23] Brian Kim (Class of 2021)
- [24] Chaerin Lim (Class of 2021)
- [25] Domnica Dzitac (Class of 2021)
- [26] Heorhii Skovorodnikov (Class of 2021)
- [27] Hochan Lee (Class of 2021)
- [28] Oorja Majgaonkar (Class of 2021)
- [29] Rashid Alzarooni (Class of 2021)
- [30] Xixuan Liu (Class of 2021)
- [31] Alia El Kattan (Class of 2020)

- [32] Anh Le Xuan Mai (Class of 2020)
- [33] Daniel Watson (Class of 2020)
- [34] Gerardo Samuel Ixcaragua (Class of 2020)
- [35] Ross Jiang (Class of 2020)
- [36] Guyu Fan (Class of 2019)
- [37] Muaz Ahmad (Class of 2019)
- [38] Paula Valentina Dozsa (Class of 2019)
- [39] Theodore Tenev (Class of 2019)
- [40] Tomasz Bachosz (Class of 2019)
- [41] Yana Chala (Class of 2019)
- [42] Angela Orozco (Class of 2018)
- [43] Asyrique Asyraf Thevendran (Class of 2018)
- [44] Dana Abu Ali (Class of 2018)
- [45] Lingliang Zhang (Class of 2017)

## Teaching Service

- [1] Ph.D. Program coordinator of the Computer Science Program, NYUAD, 2016-2019.
- [2] Ph.D. student coordinator for the Center for Computational Learning Systems (CCLS) 2012-2014.

## Guest Lectures

- [1] **Habash, Nizar.** A Palestinian Family's Exodus and Odyssey. ACS-UH 2614X Colonization of Palestine, NYU Abu Dhabi, (Spring 2021).
- [2] **Habash, Nizar.** Phonetics, CADT-UH 1004Q Rhythm, NYU Abu Dhabi, (Fall 2018).
- [3] **Habash, Nizar.** Arabic Language Technologies, CDAD-UH 1027E 5000 Years of Notable Lives, NYU Abu Dhabi, (Fall 2018).
- [4] **Habash, Nizar.** On Language Construction, COREAAD 19: Communication and Technology, NYU Abu Dhabi, (Fall 2018, Fall 2017, Fall 2016, Fall 2015).
- [5] **Habash, Nizar.** The Palisra Project, COREAAD 50W: Identity and Object, NYU Abu Dhabi, (Fall 2016, Fall 2015).
- [6] **Habash, Nizar.** Stories, ARABL-AD 218: Colloquial Arabic: Levantine Dialect 1, NYU Abu Dhabi, (Fall 2018, Fall 2016, Fall 2015).
- [7] **Habash, Nizar.** Introduction to Machine Translation, COMS 4701: Artificial Intelligence, Columbia University, **Fall 2014 (December 4, 2014).**
- [8] **Habash, Nizar.** Introduction to Arabic Natural Language Processing, CS 4705: Introduction to Natural Language Processing, Columbia University, **Fall 2013, Fall 2012, Fall 2011.**
- [9] **Habash, Nizar.** Machine Translation. Mock Class, New York University Abu Dhabi, **Spring 2013.**
- [10] **Habash, Nizar.** Machine Translation: Challenges, Approaches and Evaluation, CS 4705: Introduction to Natural Language Processing, Columbia University, **Fall 2010, Fall 2008-2004.**
- [11] **Habash, Nizar.** Machine Translation: Challenges and Approaches, CMSC 421: Introduction to Artificial Intelligence, University of Maryland College Park, **Fall 2003.**
- [12] **Habash, Nizar.** Generation Heavy Hybrid Machine Translation, CMSC 723/LING 645: Introduction to Computational Linguistics, University of Maryland College Park, **Spring 2003.**
- [13] **Habash, Nizar.** Computers and Writing Systems, HONR 279I: History of the Alphabets: Languages and their Scripts, University of Maryland College Park, **Spring 2003.**
- [14] **Habash, Nizar.** Natural Language Generation, CMSC 828: Advanced Natural Language Processing: Theory and Practice, University of Maryland College Park, **Spring 2002.**

## Short Courses and Tutorials

- [1] **Habash, Nizar**, Salam Khalifa, Ossama Obeid, Dima Taji, Go Inoue and Bashar ALhafni. Text Analysis of Arabic. A three-day short course, 10 attendees, Winter Institute in Digital Humanities (WIDH 2020), New York University Abu Dhabi, January 19-22, 2020.
- [2] **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
- [3] **Habash, Nizar**. Introduction to Arabic Natural Language Processing.
  - A three-hour tutorial. The Artificial Intelligence Program, Oxford University, Online, June 8, 2021.
  - A three-hour tutorial, 40 attendees. The Artificial Intelligence Program, Oxford University, Dubai, February 24, 2020.
  - A three-hour tutorial, 40 attendees. Arabic Natural Language Processing. The Artificial Intelligence Program, Oxford University, Dubai, February 19, 2019.
  - A one-hour tutorial, 35 attendees. In Arab Women in Computing Conference. Beirut, 2017.
  - A four-hour tutorial, 25 attendees, Carnegie Mellon Qatar, Doha, Qatar, 2010.
  - 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. Johns Hopkins University Summer School on Human Language Technology, Baltimore, 2006.
  - A three-hour tutorial, 40 attendees. BBN Technologies, Boston, Massachusetts, 2005.
  - A three-hour tutorial, 20 attendees. The National Cryptologic Museum, 2005.
  - A three-hour tutorial, 50 attendees. Johns Hopkins University Summer Workshop, Baltimore, Maryland, 2005.
  - 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.
- [4] 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.
- [5] **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.
- [6] 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.
- [7] 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.

- [8] **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.

## 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] MIT Technology Review Arabic Edition – Advisory Board Member (starting May 2018)
- [2] 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/> (2017-2018)
- [3] Member of the Editorial Board of Computer Speech and Language (2016-2017)
- [4] Member of the standing review committee of Transactions of the Association for Computational Linguistics (TACL) (2014-2016)
- [5] Member of the editorial board of Computational Linguistics Journal (2012-2014)
- [6] Co-editor with Hany Hassan for a special issue of MT journal on Arabic Machine Translation, 2012.
- [7] Member of the editorial board of the Journal for Machine Translation (2007-2015)

### Journal Reviewing

- [1] **BJMES** British Journal of Middle Eastern Studies: 2020
- [2] **CL** Journal of Computational Linguistics: 2014, 2013, 2012, 2011, 2010, 2006, 2003
- [3] **CSL** Journal of Computer Speech & Language: 2019, 2016, 2014, 2013
- [4] **FTIR** Foundations and Trends in IR: 2014
- [5] **JAIR** Journal of Artificial Intelligence Research: 2011
- [6] **JIPM** Journal of Information Processing and Management: 2017, 2016
- [7] **JJCIT** Jordanian Journal of Computers and Information Technology: 2018
- [8] **JKBS** Journal of Knowledge-Based Systems: 2016
- [9] **JKSU** Journal of King Saud University: 2018, 2016
- [10] **JPRR** Journal of Pattern Recognition Research: 2011
- [11] **LRE** Language Resources and Evaluation Journal: 2015, 2014
- [12] **MTJ** Machine Translation Journal: 2014, 2013, 2012, 2011, 2010, 2009, 2008, 2007
- [13] **NLE** Journal of Natural Language Engineering: 2020, 2019, 2014, 2013, 2012, 2011, 2007, 2004
- [14] **PAL** Perspectives on Arabic Linguistics: 2021, 2020, 2014, 2013
- [15] **SCI NLP in AI** SCI Natural Language Processing in Artificial Intelligence: 2021.
- [16] **TACL** Transactions of the Association for Computational Linguistics: 2015, 2013
- [17] **TALIP** ACM Transaction on Asian Languages Information Processing: 2014, 2010, 2004
- [18] **TASLP** IEEE Transactions on Audio, Speech and Language Processing: 2007, 2006, 2005, 2004

## 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 and Session Chairs

- [1] **\*SEM** Conference on Lexical and Computational Semantics: 2013
- [2] **AAAI** Conference of the Association for the Advancement of Artificial Intelligence: 2014
- [3] **ACL** Conference of the Association for Computational Linguistics: 2020, 2019, 2016, 2013, 2012, 2008, 2007, 2006, 2002
- [4] **ACLING** Arabic Computational Linguistics Conference: 2021, 2018, 2017
- [5] **ALTIC** International Conference Arabic Language Technology: 2011
- [6] **AMTA** Conference of the Association for Machine Translation in the Americas: 2014, 2012, 2010, 2008, 2006
- [7] **ArabWIC** International Conference of Arab Women in Computing: 2019, 2017
- [8] **ARR** ACL Rolling Review: 2021
- [9] **CAASL** Workshop on Computational Approaches to Arabic Script Languages: 2009, 2004
- [10] **CASL** Workshop on Computational Approaches to Semitic Languages: 2007, 2005
- [11] **CITALA** International Conference on the Arabic Language Processing: 2007
- [12] **COLING** Conference on Computational Linguistics: 2018, 2008, 2004,
- [13] **CONLL** Conference on Computational Natural Language Learning: 2013
- [14] **EACL** Conference of the European ACL: 2017, 2016, 2014, 2012, 2006
- [15] **EAMT** Conference of the European Association for Machine Translation: 2020, 2018
- [16] **EMNLP** Conference on Empirical Methods for Natural Language Processing: 2020, 2019, 2018, 2017, 2015, 2009, 2008, 2007
- [17] **Forum** Conference of the Forum for Arabic Linguistics: 2015
- [18] **iArabic** Palestinian Symposium on Computational Linguistics and Arabic Content: 2014
- [19] **INLG** International Natural Language Generation Conference: 2002
- [20] **Interspeech** Conference of the International Speech Communication Association: 2012
- [21] **IWPT** International Conference on Parsing Technologies: 2013
- [22] **LAW** Linguistic Annotations Workshop: 2017, 2015.
- [23] **LREC** Language Resources and Evaluation Conference: 2020, 2018, 2016, 2014, 2012, 2008, 2006
- [24] **MEDAR** Conference on Mediterranean Arabic Language and Speech Technology: 2009
- [25] **MTSummit** Machine Translation Summit: 2019, 2013, 2011, 2007
- [26] **NAACL** Conference of the North American ACL: 2019, 2018, 2015, 2013, 2012, 2007, 2006, 2001
- [27] **OSACT** Workshop on Free/Open-Source Arabic Corpora and Corpora Processing: 2018, 2016, 2014
- [28] **SeMat** Workshop on Semitic Machine Translation: 2016
- [29] **SemEval** International Workshop on Semantic Evaluation: 2017, 2015
- [30] **SIGDIAL** Conference of the Special Interest Group on Dialogue (SIGDIAL): 2018
- [31] **SIGMORPHON** Workshop of the Special Interest Group on Computational Morphology and Phonology: 2021, 2019, 2018, 2016
- [32] **SRW** Student Research Workshop at ACL: 2016
- [33] **TMI** Theoretical and Methodological Issues in Machine Translation: 2007
- [34] **UDW** Universal Dependencies Workshop: 2017
- [35] **VarDial** Workshop on NLP for Similar Languages, Varieties and Dialects: 2021, 2019, 2018, 2017, 2016, 2015
- [36] **WANLP** Workshop on Arabic Natural Language Processing: 2021, 2020, 2019, 2017, 2015, 2014
- [37] **WMT** Conference (Previously Workshop) on Statistical Machine Translation: 2021, 2019, 2018, 2017, 2016, 2015, 2011, 2010

[38] **WNUT** Workshop on Noisy User-generated Text: 2019

### **Professional Organizations**

- [1] Secretary of the board of the International Association for Machine Translation (IAMT) (2009-2018)
- [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] Chair of the Local Organizing Committee for the Conference on Empirical Methods for Natural Language Processing (EMNLP 2022) in Abu Dhabi, UAE, 2022.
- [2] Founding President of the Special Interest Group for Arabic NLP (SIGARAB), 2022.
- [3] Elected General Chair of the Workshop for Arabic Natural Language Processing (2021).
- [4] Co-organizer of the NADI Shared Task on Nuanced Dialect Identification (2020, 2021).
- [5] EMNLP 2020 Best paper selection committee.
- [6] Area Chair in EMNLP 2019 ("Phonology, Morphology and Word Segmentation").
- [7] Session Chair in NAACL 2019 and MTSummit 2019.
- [8] Ex-General/Advisor of the Workshop for Arabic Natural Language Processing (2019).
- [9] General Chair of the Workshop for Arabic Natural Language Processing (2017).
- [10] Co-organizer of the Workshop on Arabic Dialect Technologies (WARDAT), Abu Dhabi, 2016
- [11] SEM Eval Steering Committee (2016)
- [12] Co-chair of the Arabic Natural Language Processing Workshop in EMNLP (2014 and 2015).
- [13] Co-organizer of the QALB Shared Task on Arabic Spelling Correction (2014, 2015).
- [14] Publication co-chair for NAACL 2012.
- [15] Workshop program co-chair for NAACL 2009.
- [16] Senior Program Committee member for NAACL 2007 (machine translation track).
- [17] Program co-chair for the conference of the Association for Machine Translation in the Americas (AMTA'06)
- [18] One of three organizers for a workshop on Computational Approaches to Semitic Languages ACL 2005.
- [19] Chair of the Workshop on Determining Interlingua Utility for Machine Translation. AMTA-2004, Washington DC, 2004.
- [20] 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 Advisory Council of the Arabic Language in the UAE. 2020-present.
- [2] Member of the Oversight committee of the UAE Artificial Intelligence Expert Group (AIEG). 2020-present.
- [3] Advisory Board Member. Islamicate Digital Humanities Network. 2020-present.
- [4] Advisory Committee Member for Humanitas.ai. 2018-present.
- [5] Member of the Qatar Computational Research Center (QCRI) Arabic Language Technology Roundtable. Doha, Qatar, January 2011.
- [6] Co-chair of the advisory committee to improve the quality of the Arabic Treebank, 2006-2008.
- [7] Member of an advisory committee to NIST MT Open Evaluation program, 2007.
- [8] Member of the machine translation team under the NSF funded MINDS workshop, 2006.

## University Service

### *New York University Abu Dhabi*

- [1] Faculty Mentor for Yasir Zaki (CS), Thomas Poetsch (CS), Mai Oudah (CS), Farah Shamout (Engineering), Tuka Alhanai (Engineering), Hanan Salam (CS).
- [2] A member of the Task Force on NYUAD's Academic Strategy (AY 20-21)
- [3] A member of the NYUAD search committee for the Dean of Arts and Humanities (AY19-20).
- [4] A member of the NYUAD Promotion and Tenure Committee (Jan 2019- now)
- [5] Chair of the CS faculty search committee for open rank position (AY19-20; AY20-21)
- [6] A member of the NYUAD Computer Science faculty review committees for tenure (AY19-20, AY18-19)
- [7] Co-chair of the CS faculty search committee (AY18-19, AY17-18 and AY16-17).
- [8] A member of a Social Science Division faculty search committee on Computational Social Scientists (2018).
- [9] Judge in the Annual New York University Abu Dhabi (NYUAD) International Hackathon for the Social Good in the Arab World. (Spring 2017, Spring 2018, Spring 2019).
- [10] Chair of the three-year tenure-track faculty review committee: 2019, 2017, 2016.
- [11] Ph.D. Program coordinator of the Computer Science Program, NYUAD, 2016-2019.
- [12] MC for the NYUAD Computer Science Open House (Fall 2016).
- [13] A member of the Core Curriculum Committee, NYUAD, Sep 1, 2016 – Dec 15, 2017.
- [14] A member of an Arts & Humanities Division faculty search committee on Digital Humanities (2016, 2017).
- [15] Mentor. Hackathon for Social Good in the Arab World. NYUAD, 2015 and 2016.
- [16] Tower of Babel Student Interest Group advisor (2015-present).
- [17] Co-director of the CS Unix Lab, 2015-2016.
- [18] A member of the Freshman Colloquium Committee, 2015-2016.
- [19] A member of the Institutional Review Board, NYUAD, April 15, 2015-Aug 31, 2016.
- [20] A member of the doctoral student selection committee, Computer Science, NYUAD January 20 2015-now.
- [21] A member of the CS faculty search committee (for the year 2014-2015)
- [22] Advisor of the Student Interest Group "Tower of Babel" (2016-present)
- [23] Led discussion group of Freshmen Class summer reading as part of Marhaba Week:  
"Parable of the Sower" - 2015, NYUAD.  
"Bamboo Stalk" - 2016, NYUAD
- [24] Participated in the NYUAD undergraduate recruitment effort at Candidate Weekend by giving introductory lectures on machine translation, interviewing candidates, or representing the computer science program (every year 2015-2018).
- [25] Judge. Sheikh Mohamed bin Zayed Scholars Program Selections, NYUAD, 2015.
- [26] Judge. AngelHack at NYUAD, 2015.

### *Columbia University*

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

## Media (Selections)

---

- [1] Khaleej Times newspaper: UAE researchers develop unique Arabic thesaurus. Jan 23, 2021. (link [here](#))
- [2] Albyan newspaper: الذكاء الاصطناعي في خدمة لغة الضاد. Jan 23, 2021. (link [here](#))

- [3] Alroeya newspaper: نيويورك أبوظبي تبسط روائع الأدب العربي بمعجم افتراضي. Jan 19, 2021. (link [here](#))
- [4] Alittihad newspaper: اللغة العربية: أين مكانها في عالم الرقمنة. Dec 24, 2020. (link [here](#)).
- [5] Albayan newspaper: تحديات رقمنة المحتوى محور ندوة "أسبوع العربية". Dec 24, 2020. (link [here](#)).
- [6] Alkhaleej newspaper: اللغة العربية والرقمنة.. "حلول لتطوير المحتوى". Dec 23, 2020. (link [here](#)).
- [7] Albayan newspaper: تقرير حالة اللغة يرسم ملامح مستقبل العربية. Dec 20, 2020. (link [here](#)).
- [8] Emirates News Agency: أعضاء المجلس الاستشاري للغة العربية: تقرير حالة اللغة العربية يرسم استراتيجيات وملامح مستقبلها. Dec 19, 2020. (link [here](#)).
- [9] MIT Technology Review in Arabic: في اليوم العالمي للغة العربية: أهم التحديات وتقنيات الذكاء الاصطناعي في معالجة. Dec 18, 2020. (link [here](#)).
- [10] Alkhaleej: الإمارات تدعم الجهود العالمية لتوحيد معايير تطبيقات الذكاء الاصطناعي. Oct 8, 2020. (link [here](#)).
- [11] The National: How do you say 'Hey Siri' in spoken Arabic? NYUAD lab is working on Gulf dialects to find the answer. June 1, 2020. (link [here](#)).
- [12] 2019 NYUAD Research Report: "Programming Prejudice". (link [here](#)).
- [13] Al Arab newspaper: تطوير تطبيقات للغة العربية يواجه تحديات تقنية كثيرة. Sep 23, 2019 (link [here](#)).
- [14] Al Sharq Al Awsat newspaper: باحثون يطورون برنامجاً للتعرف على 25 لهجة عربية. July 11, 2019. (link [here](#)).
- [15] Interviewed as a guest on the BBC Arabic talk show Dunyana discussing challenges of Arabic language computational processing. Aired August 18, 2017.
- [16] Interviewed for an article on Wamda.com. "[Exploring the frontiers of Arabic natural language processing](#)." May 2016.
- [17] Interview with The Gazelle, the NYUAD Magazine. "[The art of artificial languages](#)". March 6, 2016.
- [18] Interviewed for an article on [machine translation for Arabic speakers \(in Arabic\)](#) on Radio Sawa Website. April 19, 2013.
- [19] Interviewed for an article on machine translation in Wired Magazine (Issue 14.12 - December 2006)
- [20] 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.

## Honors

---

- [1] قائمة إم أي تي : "30 leading arab experts in AI" Listed in MIT Technology Review Arabia list of <https://technologyreview.ae/ai-person/> ريفيو لأبرز خبراء الذكاء الاصطناعي العرب 2022. <https://technologyreview.ae/ai-person/>
- [2] Member of the Mohammed bin Rashed Academy of Scientists (MBRAS), 2021- present.
- [3] WOLDA Logo Award of Excellence in the Logo Redesign category for CAMEL Lab's new logo. Dec 20, 2020. (link [here](#)).
- [4] Appointed Member of the Advisory Council of the Arabic Language in the UAE. 2020-present.
- [5] Selected by Dubai Future Foundation to be one of the judges for 'Innovators Under 35 MENA', 2019 and 2020.
- [6] Mentored the team that was selected to represent NYUAD for the international Hult Prize competition (2016).
- [7] 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.
- [8] 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?
- [9] Received three Google Research Awards (in 2009, 2010 and 2012).
- [10] 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.
- [11] Graduate Student Service Award, University of Maryland College Park, 2003.
- [12] The Samuel N. Alexander Fellowship, Association for Computing Machinery, Washington D.C. Chapter, 2001.

- [13] The Phi Kappa Phi Award of Excellence, Phi Kappa Phi, 1997/1998.
- [14] The College of Engineering and Technology Outstanding Scholar Award, Old Dominion University, 1997.
- [15] The Faculty Award in Computer Engineering, College of Engineering and Technology, Old Dominion University, 1997.
- [16] The Outstanding Individualized Study Student Award, College of Arts and Letters, Old Dominion University, 1997.
- [17] The Award of Academic Excellence, Academic Honors Program Old Dominion University, 1997.
- [18] The Meghan O'Connor Award for Academic Achievement and Community Service, Academic Honors Program, Old Dominion University, 1997.
- [19] Undergraduate Research Award, Academic Honors Program, Old Dominion University, 1994.

## **References**

---

Available upon request.