

Houda Bouamor, PhD

Associate Area Head of Information Systems

Associate Teaching Professor of Information Systems

Work address (Qatar): Carnegie Mellon University in Qatar
Education City, P.O Box 24866, Doha, Qatar

Work address (USA): Carnegie Mellon University
Dietrich College of Humanities and Social Sciences
Pittsburgh, PA, USA

Phone: (+974) 33.15.98.56

Email: hbouamor@cmu.edu

Profile

- PhD. in Computer Science and Natural Language Processing (NLP)
- Experienced in:
 - Natural Language Processing, applied machine learning and deep learning
 - Arabic Dialect Processing
 - Collecting, building and maintaining corpora to train and validate NLP models
 - Large Language Models and Generative AI
 - Working with unstructured data both formal and informal genres
 - Handling languages with various morphological granularity and syntactic structure e.g. Arabic, French, English
 - Supervising student research activities of different levels: junior and senior thesis and projects.

Research Interests

- Applied machine learning and Deep learning
- Supervised and unsupervised learning methods
- NLP: monolingual and multilingual processing, natural language processing for morphologically complex languages, dialectal Arabic processing, low-resource languages processing, named entity recognition and translation, statistical and neural machine translation, machine translation evaluation and confidence estimation, paraphrasing, dialect identification, user-generated content processing and social media analysis, sentiment analysis, eye-tracking for NLP, automatic spelling correction, financial data processing, natural language generation, automatic essay scoring.
- Data Science: Data collection, Data processing, Data extraction and visualization, Data analysis, etc. (with a focus on textual data)

Key Experience

- Researcher: Natural Language Processing
- Project Manager: teams of 2-10 people
- Professor: Principles of Computing, Fundamental of programming, Data structures and Machine learning
- Advising: Masters and Bachelor end-of-year thesis

Notable Projects

- Arabic Automatic Essay scoring
- Deep Learning for Machine Translation
- Deep learning models for classification
- Informal text processing and translation
- Statistical/neural machine translation
- Building a multi-parallel corpus and Lexicon for Dialectal Arabic Processing
- Arabic Named Entity Translation Lexicon

Key Achievements

- (2023-2026) Grant of 600K USD funding for a research project on Automatic Scoring of Essays
- (2020-2024) Grant of 650K USD funding for a research project on Qatari Dialect Collection and Processing
- (2017-2020) Grant of 1 million USD funding for a research project on Dialectal Arabic Processing
- Top-tier journal and conference publications (ACM, ACL, EMNLP, EACL, NAACL, COLING)
- Anita Borg Institute scholarship to attend the Grace Hopper Conference 2013
- Best paper award in the 6th International Joint Conference on Natural Language Processing (IJCNLP 2013)
- Full PhD scholarship from the Ministry of Higher Education and Research in France

Education

2008 – 2012	PhD in Computer Science <i>University Paris-Sud 11, Orsay, France</i> Major in Computational Linguistics
2007 – 2008	Master of Science in Computer Science <i>University Paris-Est Marne-La-Vallée, France</i> Major in Scientific and Technical Information Processing
2003 – 2007	Bachelor of Computer Science <i>School of Business, University of Manouba, Tunis, Tunisia</i> Major in Computer Science Applied to Business Management

Research and Work Experience

May. 2019–	Researcher affiliated with the Camellab https://nyuad.nyu.edu/en/research/faculty-labs-and-projects/computational-approaches-to-modeling-language-lab/researchers.html New York University Abu Dhabi <i>Arabic NLP</i>
May. 2018– Feb. 2019	Chief Research Scientist and NLP Expert Fortia Financial Solutions <i>Financial Data Analysis and Processing, Financial Data Generation</i>
Oct. 2014 – Aug. 2018	Senior Researcher Carnegie Mellon University in Qatar <i>Arabic Natural Language Processing, Colloquial Arabic processing, Statistical Machine Translation, Sentiment Analysis, Automatic Text Correction, Machine Translation Evaluation, Quality Estimation, Arabic diacritization</i>
Sep. 2012 – Oct. 2014	Postdoctoral Research Associate Carnegie Mellon University in Qatar <i>Arabic Natural Language Processing, Colloquial Arabic processing, Statistical Machine Translation, Sentiment Analysis</i>
2008 – 2012	Research Assistant LIMSI-CNRS, Orsay, France Information, Language, writtEn and Signed Group (ILES) <i>PhD thesis title: Technique Combination for Paraphrase Alignment : Contribution to Pinpointing Sub-sentential Paraphrases</i> <i>Supervisors: Aurélien Max and Anne Vilnat</i>
Apr. 2007 – Sep. 2008	Research Intern CEA-LIST , France Multilingual Multimedia knowledge engineering lab (LIC2M) <i>Master thesis title: Mining a Multilingual Geographical Gazetteer from the Web</i> <i>Supervisors: Adrian Popescu, Pierre-Alain Moellic</i>
June – Aug. 2007	Research Intern LLACAN-CNRS , France <i>Topic: Arabic Natural Language Processing: Interactive Annotation of Arabic Corpus</i>

Teaching Experience

July 2022 –	Associate Teaching Professor <i>Courses: Application Design and Development (67-272), Natural Language Processing with Python (67-212), Practical Data Science (67-364), Capstone course in Data Science (67-426)</i> Carnegie Mellon University in Qatar
March 2019 – June 2022	Assistant Teaching Professor <i>Courses: Application Design and Development (67-272), Introduction to Data Analysis (67-335), Data Analysis for social commerce (99-520-Z), Natural Language Processing with Python (67-212), Practical Data Science (67-364)</i> Carnegie Mellon University in Qatar
August 2015 – August 2018	Visiting Assistant Teaching Professor <i>Courses: Principles of Computing (15-110), Fundamentals of Programming and Computer Science (15-112), Introduction to Data Structures (15-121), Introduction to Machine Learning (10-601)</i> Carnegie Mellon University in Qatar
Summer 2016	Visiting Assistant Teaching Professor <i>Course: Introduction to Data Structures (15-121)</i> Carnegie Mellon University, Pittsburgh
Summer 2015	Instructor <i>Courses: Fundamentals of Programming and Computer Science(15-112)</i> Carnegie Mellon University in Qatar
Fall 2013	Guest Lecture <i>Lecture: Introduction to Text Processing: POS Tagging</i> Carnegie Mellon University in Qatar
2011 – 2012	Teaching Assistant (ATER) <i>Courses: Algorithms and, C and Java Programming Languages</i> Polytech' Paris Sud & Paris Sud University, UFR des Sciences
2008 – 2011	Teaching Assistant (Monitrice) <i>Courses: Algorithms, Information System Design, Databases, Java Programming language</i> Institut Universitaire de Technologies (IUT) d'Orsay, University Paris-Sud 11

Publications

Patents

1. **Houda Bouamor**, Sira Ferradans, Guillaume Hubert and Abderrahim Ait-Azzi, "Generation of Natural Language Text from Structured Data using a Fusion Model". 2021. US/163613009. .

The following publications are also available from my Google citation page.

Refereed Journal Publications

1. Kareem Darwish, Nizar Habash, Mourad Abbas, Hend Al-Khalifa, Huseein 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". 2021. *Communications of the ACM* 64 (4), 72-81
2. **Houda Bouamor**, Aurélien Max, Anne Vilnat, "Multi-technique paraphrase alignment: A contribution to pinpointing sub-sentential paraphrases", *ACM Transactions on Intelligent Systems and Technology (Special Issue on Paraphrasing)*. 2013. Volume 4. Issue 3.
3. **Houda Bouamor**, Aurélien Max, Anne Vilnat, "Étude bilingue de l'acquisition et de la validation automatiques de paraphrases sous-phrastiques" *Traitement Automatique des Langues*, 2012, volume 53.1
English title: "A bilingual study of sub-sentential paraphrase automatic acquisition and validation."
4. Camille Dutrey, **Houda Bouamor**, Delphine Bernhard, Aurélien Max, "Local modifications and paraphrases in Wikipedia's revision history. " *SEPLN journal (Revista de Procesamiento del Lenguaje Natural)*, volume 46:51-58.

Refereed International Conference and Workshop Publications

1. **Houda Bouamor**, Gabriela Gongora-Svartzman, Larry Heimann, and Shihong Huang. "Evaluating GenAI's Effectiveness for Students with Varied Programming Backgrounds in a Software Development Course". SIGCSE TS 2025. Proceedings of the 56th ACM Technical Symposium on Computer Science Education. Pittsburgh, PA, USA: 2025
2. Malika Dikshit, **Houda Bouamor** and Nizar Habash. "Investigating Gender Bias in STEM Job Advertisements". GeBNLP 2024. Proceedings of the 5th Workshop on Gender Bias in Natural Language Processing (GeBNLP). Bangkok, Thailand: 2024
3. Kurt Micallef, Nizar Habash, Claudia Borg, Fadhl Eryani and **Houda Bouamor**. "Cross-Lingual Transfer from Related Languages: Treating Low-Resource Maltese as Multilingual Code-Switching". EACL 2024. Proceedings of the 18th Conference of the European Chapter of the Association for Computational Linguistics (EACL). San Giljan, Malta: 2024
4. Bashar Alhafni, Sarah Al-Towaity, Ziyad Fawzy, Fatema Nassar, Fadhl Eryani, **Houda Bouamor** and Nizar Habash. "Exploiting Dialect Identification in Automatic Dialectal Text Normalization". ArabicNLP 2024. Proceedings of the Second Arabic Natural Language Processing Conference (ArabicNLP 2024). Bangkok, Thailand: 2024
5. Bashar Alhafni, Sarah Al-Towaity, Ziyad Fawzy, Fatema Nassar, Fadhl Eryani, **Houda Bouamor** and Nizar Habash. "Exploiting Dialect Identification in Automatic Dialectal Text Normalization". ArabicNLP 2024. Proceedings of the Second Arabic Natural Language Processing Conference (ArabicNLP 2024). Bangkok, Thailand: 2024

6. May Bashendy, Salam Albatarni, Sohaila Eltanbouly, EMAN Zahran, Hamdo Elhuseyin, Tamer Elsayed, Walid Massoud and **Houda Bouamor**. "QAES: First Publicly-Available Trait-Specific Annotations for Automated Scoring of Arabic Essays". ArabicNLP 2024. Proceedings of the Second Arabic Natural Language Processing Conference (ArabicNLP 2024). Bangkok, Thailand: 2024
7. Sanad Malaysha, Mo El-Haj, Saad Ezzini, Mohammed Khalilia, Mustafa Jarrar, Sultan Al-mujaiwel, Ismail Berrada and **Houda Bouamor**. "AraFinNLP 2024: The First Arabic Financial NLP Shared Task". ArabicNLP 2024. Proceedings of the Second Arabic Natural Language Processing Conference (ArabicNLP 2024). Bangkok, Thailand: 2024
8. Muhammad Abdul-Mageed, Amr Keleg, AbdelRahim A. Elmadany, Chiyu Zhang, Injy Hamed, Walid Magdy, **Houda Bouamor** and Nizar Habash "NADI 2024: The Fifth Nuanced Arabic Dialect Identification Shared Task". ArabicNLP 2024. Proceedings of the Second Arabic Natural Language Processing Conference (ArabicNLP 2024). Bangkok, Thailand: 2024
9. Wajdi Zaghoulani, Mustafa Jarrar, Nizar Habash, **Houda Bouamor**, Imed Zitouni, Mona T. Diab, Samhaa R. El-Beltagy and Muhammed AbuOdeh. "The FIGNEWS Shared Task on News Media Narratives". ArabicNLP 2024. Proceedings of the Second Arabic Natural Language Processing Conference (ArabicNLP 2024). Bangkok, Thailand: 2024
10. **Houda Bouamor**, Juan Pino and Kalika Bali. "Findings of the Association for Computational Linguistics: EMNLP 2023". EMNLP 2023. The 2023 Conference on Empirical Methods in Natural Language Processing. Singapore, Singapore: 2023
11. Hasan Alkheder, **Houda Bouamor**, Nizar Habash, Ahmet Zengin. "Benchmarking Dialectal Arabic-Turkish Machine Translation". MT Summit 2023. Machine Translation Summit XIX, Vol. 1: Research Track. Macau SAR, China: 2023
12. Kurt Micallef, Fadhl Eryani, Nizar Habash, **Houda Bouamor** and Claudia Borg. "Exploring the Impact of Transliteration on NLP Performance for Low-Resource Languages: The Case of Maltese and Arabic". CAWL 2023. The Workshop on Computation and Written Language (CAWL 2023). Toronto, Canada: 2023
13. **Houda Bouamor**, Hend Al-Khalifa, Kareem Darwish, Owen Rambow, Fethi Bougares, Ahmed Abdelali, Nadi Tomeh, Salam Khalifa, Wajdi Zaghoulani. WANLP 2022. "Proceedings of the Seventh Arabic Natural Language Processing Workshop (WANLP)". The Seventh Arabic Natural Language Processing Workshop (WANLP). Abu Dhabi, UAE: 2022
14. Bashar Al-Hafni, Nizar Habash, and **Houda Bouamor**. "User-Centric Gender Rewriting". NAACL 2022. The 2022 Conference of the North American Chapter of the Association for Computational Linguistics - Human Language Technologies (NAACL-HLT 2022). Seattle, USA: 2022
15. Bashar Al-Hafni, **Houda Bouamor**, and Nizar Habash. "The Arabic Parallel Gender Corpus 2.0: Extensions and Analyses ". LREC 2022. 13th edition of the Language Resources and Evaluation Conference. Marseille, France: 2022.

16. Nurpeiis Baimukan, **Houda Bouamor**, Nizar Habash. "Aggregating Hierarchical Dialectal Data for Arabic Dialect Classification". *LREC 2020. 13th edition of the Language Resources and Evaluation Conference*. Marseille, France: 2022.
17. Samir Abdaljalil and **Houda Bouamor**. "An Exploration of Automatic Text Summarization of Financial Reports". *FinNLP 2021. The 3rd Workshop on Financial Technology and Natural Language Processing (FinNLP)*. *Best paper Award*
18. Go Inoue, Bashar Alhafni, Nurpeiis Baimukan, **Houda Bouamor**, and Nizar Habash. "The Interplay of Variant, Size, and Task Type in Arabic Pre-trained Language Models". *WANLP 2021. Proceedings of the Sixth Arabic Natural Language Processing Workshop*
19. Muhammad Abdul-Mageed, Chiyu Zhang, AbdelRahim Elmadany, **Houda Bouamor**, Nizar Habash. "NADI 2021: The Second Nuanced Arabic Dialect Identification Shared Task". *Proceedings of the Sixth Arabic Natural Language Processing Workshop*
20. Abdullah Shaar, Andrew Edward, **Houda Bouamor**, and Savanid Vatanasakdakul. "Measuring the Impact of COVID-19 on the Shipping in the GCC countries". *ACIS 2020. Proceedings of The Australasian Conferences on Information Systems*
21. Bashar Alhafni, Nizar Habash and **Houda Bouamor**. "Gender-Aware Reinflection using Linguistically Enhanced Neural Models". *GeBNLP 2020. The 2nd Workshop on Gender Bias in Natural Language Processing, co-located with COLING 2020, Barcelona, Spain: 2020*.
22. Muhammad Abdul-Mageed, Chiyu Zhang, **Houda Bouamor**, and Nizar Habash. "NADI 2020: The First Nuanced Arabic Dialect Identification Shared Task". *WANLP 2020. The Fifth Arabic Natural Language Processing Workshop, co-located with COLING 2020. Barcelona, Spain, 2020*.
23. Fadl Eryani, Nizar Habash, **Houda Bouamor**, and Salam Khalifa. "A Spelling Correction Corpus for Multiple Arabic Dialects". *LREC 2020. 12th edition of the Language Resources and Evaluation Conference*. Marseille, France: 2020.
24. **Houda Bouamor**, Sabit Hassan and Nizar Habash, "The MADAR Shared Task on Arabic Fine-Grained Dialect Identification". *WANLP 2019. The Fourth Arabic Natural Language Processing Workshop*. Florence, Italy, 2019.
25. Nizar Habash, **Houda Bouamor** and Christine Chung, "Gender Aware Arabic Natural Language Processing: First Person Gender Identification and Reinflection". *GeBNLP 2019. The 1st ACL Workshop on Gender Bias for Natural Language Processing*, Florence, Italy: 2019.
26. Abderrahim Ait-Azzi, **Houda Bouamor**, and Sira Ferradans. "The FinSBD-2019 Shared Task: Sentence Boundary Detection in PDF Noisy Text in the Financial Domain". *Proceedings of the First Workshop on Financial Technology and Natural Language Processing colocated with IJCAI 2019*. Macao, China: 2019.
27. Alexander Erdmann, Salam Khalifa, Mai Oudah, Nizar Habash and **Houda Bouamor**, "A Little Linguistics Goes a Long Way: Unsupervised Segmentation with Limited Language Specific Guidance". *SIGMORPHON 2019. The 16th ACL SIGMORPHON Workshop on Computational Research in Phonetics, Phonology, and Morphology*, Florence, Italy: 2019.

28. Ossama Obeid, Mohammad Salameh, **Houda Bouamor**, and Nizar Habash, "ADIDA: Automatic Dialect Identification for Arabic". *NAACL 2019 (demo Session). The Conference of the North American Association for Computational Linguistics*, Minneapolis, USA: 2019.
29. Abderrahim Ait-Azzi and **Houda Bouamor**, "Fortia₁ @ the NTCIR-14 FinNum Task: Enriched Sequence Labeling for Numeral Classification". *NTCIR-14 2019. The 14th NTCIR (2018-2019) Evaluation of Information Access Technologies (To appear)*, Tokyo, Japan: 2019.
30. Mohammad Salameh, **Houda Bouamor** and Nizar Habash, "Fine-Grained Arabic Dialect Identification". *COLING 2018. The 27th International Conference on Computational Linguistics*, Santa Fe, New Mexico, USA: 2018.
31. **Houda Bouamor** and Hassan Sajjad. "H2@BUCC18: Parallel Sentence Extraction from Comparable Corpora Using Multilingual Sentence Embeddings". *BUCC 2018. 11th workshop on Building and Using Comparable Corpora*, Miyazaki, Japan. 2018
32. **Houda Bouamor**, Nizar Habash, Mohammad Salameh, Wajdi Zaghoulani, Owen Rambow, Dana Abdulrahim, Ossama Obeid, Salam Khalifa, Fadhl Eryani, Alexander Erdmann and Kemal Oflazer. "The MADAR Arabic Dialect Corpus and Lexicon". *LREC 2018. 11th edition of the Language Resources and Evaluation Conference*. Miyazaki, Japan: 2018.
33. Ossama Obeid, Salam Khalifa, Nizar Habash, **Houda Bouamor**, Wajdi Zaghoulani and Kemal Oflazer. "MADARi: A Web Interface for Joint Arabic Morphological Annotation and Spelling Correction". *LREC 2018. 11th edition of the Language Resources and Evaluation Conference*. Miyazaki, Japan: 2018.
34. Nizar Habash, Salam Khalifa, Fadhl Eryani, Owen Rambow, Dana Abdulrahim, Alexander Erdmann, Reem Faraj, Wajdi Zaghoulani, **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". *LREC 2018. 11th edition of the Language Resources and Evaluation Conference*. Miyazaki, Japan: 2018.
35. Alexander Erdmann, Nizar Habash, Dima Taji and **Houda Bouamor**. "Low Resourced Machine Translation via Morpho-syntactic Modeling: The Case of Dialectal Arabic". *MT Summit 2017. MT Summit XVI, vol.1: Research Track*, Nagoya, Japan: 2017.
36. Francisco Guzmán, **Houda Bouamor**, Ramy Baly and Nizar Habash. "Machine Translation Evaluation for Arabic using Morphologically-enriched Embeddings". *COLING 2016. The 26th International Conference on Computational Linguistics*, Osaka, Japan: 2016.
37. Wajdi Zaghoulani, Nizar Habash, **Houda Bouamor**, Ossama Obeid, Sawsan Alqahtani, Mona Diab and Kemal Oflazer. "Filtering Dialectal Arabic Text in Two Large Scale Annotation Projects". *W-NUT 2016. The 2nd Workshop on Noisy User-generated Text*. Osaka, Japan: 2016
38. Wajdi Zaghoulani, Abdelati Hawwari, Sawsan Alqahtani, **Houda Bouamor**, Mahmoud Ghoneim, Mona Diab and Kemal Oflazer. "Using Ambiguity Detection to Streamline Linguistic Annotation". *CL4LC. Workshop on Computational Linguistics for Linguistic Complexity. Colocated with CLOING 2016*. Osaka, Japan: 2016

39. Hassan Sajjad, Francisco Guzmàn, Nadir Durrani, Ahmed Abdelali, **Houda Bouamor**, Irina Temnikova and Stephan Vogel. "Eyes Don't Lie: Predicting Machine Translation Quality Using Eye Movement". *NAACL HLT 2016. The 15th Annual Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, San Diego, CA, USA: 2016.
40. Wajdi Zaghoulani, **Houda Bouamor**, Abdelati Hawwari, Mona Diab, Ossama Obeid, Mahmoud Ghoneim, Sawsan Alqahtani and Kemal Oflazer. *Large Scale Arabic Diacritized Corpus: Guidelines and Framework*. *LREC 2016*. Portoroz, Slovenia: 2016.
41. Salam Khalifa, **Houda Bouamor** and Nizar Habash. *DALILA: The Dialectal Arabic Linguistic Learning Assistant*. *LREC 2016*. Portoroz, Slovenia: 2016
42. Ossama Obeid, **Houda Bouamor**, Wajdi Zaghoulani, Mahmoud Ghoneim, Abdelati Hawwari, Mona Diab and Kemal Oflazer. *MANDIAC: A Web-based Annotation System For Manual Arabic Diacritization*. *OSACT2 2016*. Portoroz, Slovenia: 2016
43. Alla Rozovskaya, **Houda Bouamor**, Nizar Habash, Wajdi Zaghoulani, Ossama Obeid, Behrang Mohit. *The Second QALB Shared Task on Automatic Text Correction for Arabic*. *ANLP 2015. Second Workshop on Arabic Natural Language Processing*. Beijing, China: 2015.
44. **Houda Bouamor**, Wajdi Zaghoulani, Mona Diab, Ossama Obeid, Kemal Oflazer, Mahmoud Ghoneim, Abdelati Hawwari. *A Pilot Study on Arabic Multi-Genre Corpus Diacritization*. *ANLP 2015. Second Workshop on Arabic Natural Language Processing*. Beijing, China: 2015.
45. **Houda Bouamor**, Hassan Sajjad, Nadir Durrani, Kemal Oflazer. *QCMUQ@QALB-2015 Shared Task: Combining Character level MT and Error-tolerant Finite-State Recognition for Arabic Spelling Correction*. *ANLP 2015. Second Workshop on Arabic Natural Language Processing*. Beijing, China: 2015.
46. Fethi Bougares and **Houda Bouamor**. *UMMU@QALB-2015 Shared Task: Character and Word level SMT pipeline for Automatic Error Correction of Arabic Text*. *ANLP 2015. Second Workshop on Arabic Natural Language Processing*. Beijing, China: 2015.
47. Wajdi Zaghoulani, Nizar Habash, **Houda Bouamor**, Alla Rozovskaya, Behrang Mohit, Kemal Oflazer, "Correction Annotation for Non-Native Arabic Texts: Guidelines and Corpus". *LAW IX 2015. The 9th Linguistic Annotation Workshop*, Denver, USA: 2015.
48. **Houda Bouamor**, Hanan Alshikhabobakr, Behrang Mohit, Kemal Oflazer, "A Human Judgement Corpus and a Metric for Arabic MT Evaluation". *EMNLP 2014. Conference on Empirical Methods in Natural Language Processing*, Doha, Qatar: 2014.
49. Serena Jeblee, Weston Feely, **Houda Bouamor**, Alon Lavie, Nizar Habash and Kemal Oflazer. "Domain and Dialect Adaptation for Machine Translation into Egyptian Arabic". *ANLP14. Workshop on Arabic Natural Language Processing at EMNLP 2014*, Doha, Qatar: 2014.
50. Serena Jeblee, **Houda Bouamor**, Wajdi Zaghoulani and Kemal Oflazer. "CMUQ at The 2014 Automatic Arabic Error Correction Shared Task". *ANLP 2014. Workshop on Arabic Natural Language Processing at EMNLP 2014*, Doha, Qatar: 2014.

51. Sabih Bin Wasi, Rukhsar Neyaz, **Houda Bouamor** and Behrang Mohit. "CMUQ@Qatar: Using Rich Lexical Features for Sentiment Analysis on Twitter". *Semeval'14. The Eighth International Workshop on Semantic Evaluation*, Dublin, Ireland: 2014.
52. Kamla Al-Mannai, Hanan Alshikhabobakr, Sabih Bin Wasi, Rukhsar Neyaz, **Houda Bouamor** and Behrang Mohit. "CMUQ-Hybrid: Sentiment Classification By Feature Engineering and Parameter Tuning". *Semeval'14. The Eighth International Workshop on Semantic Evaluation*, Dublin, Ireland: 2014.
53. **Houda Bouamor**, Nizar Habash, Kemal Oflazer, "A Multidialectal Parallel Corpus of Arabic" *LREC 2014. The 9th international conference on Language Resources and Evaluation*, Reykjavik, Iceland: 2014.
54. Ahmed Salama, **Houda Bouamor**, Behrang Mohit, Kemal Oflazer, "YouDACC: the Youtube Dialectal Arabic Commentary Corpus " *LREC 2014. The 9th international conference on Language Resources and Evaluation*, Reykjavik, Iceland: 2014.
55. **Houda Bouamor**, Behrang Mohit, Kemal Oflazer, "SuMT: A Framework of Summarization and MT" *IJCNLP 2013. The 6th International Joint Conference on Natural Language Processing*, Nagoya, Japan: 2013. (**Best Paper Award**)
56. Mahmoud Azab, **Houda Bouamor**, Behrang Mohit, Kemal Oflazer, " Dudley North visits North London: Learning When to Transliterate to Arabic" *NAACL-HLT 2013. Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies* , Atlanta, USA: 2013.
57. Aurélien Max, **Houda Bouamor**, Anne Vilnat, " Generalizing Sub-sentential Paraphrase Acquisition across Original Signal of Text Pairs" *EMNLP-CONLL 2012. Conference on Empirical Methods in Natural Language Processing and Natural Language Learning*, Jeju, Korea: 2012.
58. **Houda Bouamor**, Aurélien Max, Anne Vilnat, "Validation of Sub-sentential Paraphrases Acquired From Parallel Monolingual Corpora" *EACL 2012. The 13th Conference of the European Chapter of the Association for computational Linguistics*, Avignon, France: 2012.
59. **Houda Bouamor**, Aurélien Max, Gabriel Illouz, Anne Vilnat, "A contrastive review of paraphrase acquisition techniques" *LREC 2012. The 8th international conference on Language Resources and Evaluation*, Istanbul, Turkey: 2012.
60. **Houda Bouamor**, Aurélien Max, Anne Vilnat, "Une étude en 3D de la paraphrase : types de corpus, langues et techniques" *TALN 2012. 19ème Conférence sur le Traitement Automatique des Langues Naturelles*, Grenoble, France: 2012.
61. **Houda Bouamor**, Aurélien Max, Gabriel Illouz, Anne Vilnat, "Validation sur le Web de reformulations locales : application à la Wikipédia" *TALN 2012. 19ème Conférence sur le Traitement Automatique des Langues Naturelles*, Grenoble, France: 2012.
62. **Houda Bouamor**, Aurélien Max, Anne Vilnat, " Monolingual Alignment by Edit Rate Computation on Sentential Paraphrase Pairs" *ACL-HLT 2011. The 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies*, Portland, Oregon, USA: 2011.

63. **Houda Bouamor**, Aurélien Max, Gabriel Illouz, Anne Vilnat, "Web-based validation for contextual targeted paraphrasing" *Text-To-Text 2011. The ACL 2011 Workshop on Monolingual Text-To-Text Generation*, Portland, Oregon, USA: 2011.
64. **Houda Bouamor**, Aurélien Max, Anne Vilnat, "Combinaison d'informations pour l'alignement monolingue" *TALN 2011. 18ème Conférence sur le Traitement Automatique des Langues Naturelles*, Montpellier, France: 2011.
65. Camille Dutrey, **Houda Bouamor**, Delphine Bernhard, Aurélien Max, Anne Vilnat, "Paraphrases et modifications locales dans l'historique des révisions de Wikipédia" *TALN 2011. 18ème Conférence sur le Traitement Automatique des Langues Naturelles*, Montpellier, France: 2011.
66. **Houda Bouamor**, Aurélien Max, Anne Vilnat, "Acquisition de paraphrases sous-phrastiques depuis des paraphrases d'énoncés" *TALN 2010. 17ème Conférence sur le Traitement Automatique des Langues Naturelles, papier court*, Montréal, Canada: 2010.
67. **Houda Bouamor**, Aurélien Max, Anne Vilnat, "Comparison of paraphrase acquisition techniques on sentential paraphrases" *IceTAL 2010. 7th International Conference on Natural Language Processing*, Reykjavik, Iceland: 2010.
68. Adrian Popescu, Gregory Grefenstette, **Houda Bouamor**, "Mining a Multilingual Geographical Gazetteer from the Web" *WI 2009. The 2009 IEEE/WIC/ACM International Conference On Web Intelligence*, Milan, Italy: 2009.
69. **Houda Bouamor**, Aurélien Max, Anne Vilnat, "Amener des utilisateurs à créer et évaluer des paraphrases par le jeu" *TALN 2009. 16ème Conférence sur le Traitement Automatique des Langues Naturelles, Session de démonstrations*, Senlis, France: 2009.
70. **Houda Bouamor**, "Extraction des connaissances à partir du Web pour la recherche des images géoréférencées" *RJCRI 2009. 4èmes Rencontres Jeunes Chercheurs en Recherche d'Informations*, Presqu'îles de Giens, France: 2009.

Thesis/Reports

Houda Bouamor, "Étude de la paraphrase sous-phrastique en traitement automatique des langues", Université Paris Sud, École doctorale d'Informatique. PhD Thesis. 2012

Camille Dutrey, **Houda Bouamor**, Delphine Bernhard, Aurélien Max. "Typologie des modifications dans les révisions de Wikipédia". LIMSI Technical Report 2011-01

Houda Bouamor, "Extraction de connaissances à partir du Web pour la recherche d'images géo-référencées sur Internet", Université Paris-Est Marne-la-Vallée. Msc Thesis. 2008

Research Grants

Current grants

- **The GAITAR fellowship Program: Exploring the use of GenAI tools in 67-272: Application Design and Development (2024)**

Larry Heimann, Houda Bouamor, Shihong Huang

Our project aims to explore the possibility of using GenAI tools to help level the playing field between students and give less-experienced students a better chance to be successful in technical classes. Upon entering college to pursue STEM fields, students encounter an uneven playing field, with disparities in their high school preparation affecting their performance in the classroom, particularly in strongly technical-focused courses. Well-prepared students with programming experience, for example, spend less time on assignments and often outperform their less-prepared peers, leading to retention challenges. In this project, we investigate the impact of generative AI in programming-oriented classes, specifically in Application Design & Development (67-272). This course comprises over 160 students across the Pittsburgh and Doha campuses, representing diverse backgrounds and levels of preparedness. This project is funded by a GAITAR fellowship.

- **TartanBotQ: Designing and Building a ChatBot for Academic Advising at CMUQ (2024)**

Houda Bouamor

TartanBotQ is a one-year research project funded by a seed fund at Carnegie Mellon University. It is built in collaboration between Carnegie Mellon University in Qatar and Qatar University. The project will start in October 2024 and will finish by September 2025. In this project, we aim to develop a chatbot that can support conversation with students at Carnegie Mellon University in Qatar in English mainly. We aim to build a contemporary conversational chatbot that is user-friendly and possesses the capabilities to simulate human conversations while covering the different questions that students ask in the four programs offered here at CMUQ: Biological Sciences, Business Administration, Computer Science, and Information Systems, in addition to the Arts and Sciences. The project was awarded a total of \$50K USD spread over one year.

- **Dimension-Specific Automated Scoring of Arabic Language Writing Proficiency (2023-2026)**

Tamer Elsayed, Houda Bouamor, Walid Massoud

The Dimension-Specific Automated Scoring of Arabic Language Writing Proficiency (AAES) is a 3-year research project funded by the Qatar National Research Fund (QNRF). It is built in collaboration between Carnegie Mellon University in Qatar and Qatar University. The project started on March 2023, and will finish by February 2026. In this project, we aim to develop the first system that automatically scores Arabic essays written by high school and first-year university students to assist their Arabic writing proficiency in four different dimensions, i.e., style, relevance, organization, and development. Our system is planned to be integrated with Qatar University Testing Center (QUTC) and Pre-University Education schools at Qatar Foundation (QF-PUE) as end-users. The project was awarded a total of \$600K USD spread over 3 years (2023-2026) by Qatar National Research Fund Award (grant NPRP14S-0402-210127).

Past Grants

- **The Qatar Linguistic Map (2020-2024)**

Zeinab Ibrahim, Houda Bouamor, Hany Fazza, Aisha Al-Sultan

The Qatar Linguistic Map (QLM) is a 3-year research project funded by the Qatar National Research Fund (QNRF). This is built in collaboration between Carnegie Mellon University in Qatar, the Doha International Family Institute (DIFI), and Georgetown University in Qatar, all of which are institutions of the Qatar Foundation. The project started in January 2020 and finished in December 2023. In QLM, we aim to explore, identify, and complete any existing linguistic corpus of Qatari dialect and track dialectal changes among different Qatari age groups. The goal is to build an interactive linguistic map of Qatari dialects that will be hosted at Qatar National Library (similar to British and US linguistic museums). This will contribute to preserving the linguistic heritage of Qatar. The project was awarded a total of \$650K USD spread over 3 years (2020-2022) by the Qatar National Research Fund Award (grant NPRP12S-0301-190189). More details on the project can be found at: <https://sites.google.com/view/qlm-cmuq/home/>

- **MADAR: Multi-Arabic Dialect Applications and Resources (2017-2020)**

Nizar Habash, Houda Bouamor, Kemal Oflazer and Owen Rambow

MADAR is a three-year joint project among the NLP Group at Carnegie Mellon University in Qatar (CMU-Q), the Computational Approaches to Modeling Language (CAMEL) Lab at New York University Abu Dhabi (NYUAD), and Columbia University. The project also involves collaborators from the University of Bahrain (UoB). The project aimed at improving dialectal Arabic processing by: (1) developing resources for Arabic Dialect modeling, including the creation of a 25-city multi-dialect lexicon and a 25-city multi-dialect parallel corpus; (2) developing machine translation systems among dialects, dialects and English, dialects and Standard Arabic; and (3) developing dialect identification systems that can work on a variety of granularity levels. The MADAR Project is the largest in scale and depth to date when it comes to working on natural language processing of Arabic dialects. The project was awarded a total of \$966,551 USD spread over 4 years (2013-2017) by Qatar National Research Fund Award (grant NPRP-7-290-1-047). More details on the project could be found at: <https://camel.abudhabi.nyu.edu/madar/>

- **QALB: The Qatar Arabic Language Bank (2012-2015)**

Kemal Oflazer, Nizar Habash, Behrang Mohit, Houda Bouamor, and Nadi Tomeh

QALB is a collaboration between Carnegie Mellon University in Qatar (CMUQ), Columbia University, University of Paris-Sud, and New York University in Abu Dhabi. During that project, the team built the QALB corpus, a large manually corrected corpus (2 million words) for a variety of Arabic texts, including texts produced by native and non-native writers, as well as machine translation output. The errors annotated in the QALB corpus include spelling, punctuation, word choice, morphology, syntax, and dialectal usage at the regional level. Furthermore, the project included research on several techniques for automatic correction of Arabic text. The QALB corpus includes commentaries written in response to Al Jazeera articles and L2 essays (text written by learners of Arabic as a Foreign Language). All of the data includes corrections of language errors in these texts by native Arabic speakers at CMUQ. The QALB Project was awarded 1 million USD spread over 3 years (2012-2015) by the Qatar National Research Fund (grant NPRP-4-1058-1-168). More details on the project could be found at: <http://nlp.qatar.cmu.edu/qalb/>.

- **OptDiac: An Optimal Diacritization Scheme for Arabic Orthographic Representation (2013-2016)**

Mona Diab, Kemal Oflazer and Houda Bouamor

OptDiac is a three-year project among the NLP Group at Carnegie Mellon University in Qatar (CMU-Q) and The George Washington University (GWU). The project aimed to find an optimal diacritization system that encodes the essential level of information to be explicitly marked in the script (OPTDIAC). Their research program investigates different hypothesized diacritization schemes from different scientific perspectives: both psycholinguistic and neurolinguistic studies coupled with computational modeling in the context of natural language processing machinery. OPTDIAC produces the optimal diacritization automatically on already existing text, and may be used by written media publishing houses as well as the internet. The project was awarded a total of \$866,551 USD spread over 3 years (2013-2016), by Qatar National Research Fund Award (grant NPRP-6-1020-1-199).

Main Authored Resources

- **The MADAR Arabic Dialect Corpus**

This is a collection of parallel sentences covering the dialects of 25 cities from the Arab World, in addition to English, French, and Modern Standard Arabic (MSA). This corpus is created by translating selected sentences from the Basic Traveling Expression Corpus (BTEC) in French and English to the different dialects. The MADAR Corpus is available to the research community under a non-commercial license at: <https://camel.abudhabi.nyu.edu/madar-shared-task-2019/>

- **The MADAR Lexicon**

This is a collection of 1,045 concepts extracted from the MADAR Corpus defined in terms of triplets of words and phrases from English, French, and MSA, along with multiple equivalent dialectal forms covering 25 cities from the Arab World. Each dialectal form includes its CODA orthography and CAPHI phonology. The MADAR Lexicon will be made available soon to the research community under a non-commercial license. The lexicon can be viewed at: <https://camel.abudhabi.nyu.edu/madar/#LexiconViewer>

- **The Arabic Dialect Twitter Corpus**

This is a collection of 2,980 Twitter user profiles from 21 different Arab countries. These Twitter user profiles reflect the way users from different regions in the Arab World tweet using their own dialects. The profiles were all manually assigned a label indicating the native dialect of the user, by a team of three annotators. For each labeled user profile, only the first 100 available tweets at collection time are kept. More details on this dataset can be found at: <https://www.aclweb.org/anthology/W19-4622.pdf>

- **MADARi: The MADAR annotation interface**

MADARi is a web-based framework consisting of a management interface that allows a lead annotator to upload documents and assign tasks to annotators and an annotation

interface that allows annotators to efficiently annotate both Standard and Dialectal Arabic texts. MADARi could be found here: <https://madari.abudhabi.nyu.edu>

- **CODA***: Unified Guidelines and Resources for Arabic Dialect Orthography
CODA* (pronounced CODA Star, as in, for any dialect) is a conventional orthography for dialectal Arabic. It is designed primarily for the purpose of developing computational models of Arabic dialects. The guidelines are available at: <https://camel-guidelines.readthedocs.io/en/latest/orthography/>
- **AL-BLEU dataset and software**
This is a human judgments dataset and an adapted metric for evaluation of Arabic machine translation. This dataset is the first of its kind for Arabic with high annotation quality. The dataset and the evaluation software are available at: <http://nlp.qatar.cmu.edu/resources/AL-BLEU/>
- **A Multidialectal Parallel Corpus of Arabic**
This corpus was the first multidialectal Arabic parallel corpus. It is a collection of 2,000 sentences in Standard Arabic, Egyptian, Tunisian, Jordanian, Palestinian and Syrian Arabic, in addition to English. Such parallel data did not exist naturally, which made this corpus a very valuable resource that has many potential applications such as Arabic dialect identification and machine translation.
- **YouDACC: the Youtube Dialectal Arabic Comment Corpus**
YOUDACC is an automatically annotated large-scale multi-dialectal Arabic corpus collected from user comments on Youtube videos. This corpus covers different groups of dialects: Egyptian, Gulf, Iraqi, Maghrebi, and Levantine.
- **The Automatic Arabic Error Correction QALB data**
A large manually corrected corpus of errors for a variety of Arabic texts, including user comments on news websites, native and non-native speaker essays, and machine translation output. The corpus is available for download at: <https://camel.abudhabi.nyu.edu/qalb-shared-task-2015/>
- **The Wikipedia Correction and Paraphrase Corpus (WiCoPaCo)**
WiCoPaCo is a corpus of natural rewritings extracted from the revision history of Wikipedia. It includes spelling corrections, reformulations, and other local text transformations, which can be of great interest for many NLP applications, such as text correction, text normalization, paraphrasing or summarization. The corpus is available at: <https://wicapaco.limsi.fr/>

Honors/Awards

- The Eberly GAITAR fellowship Program: Exploring the use of GenAI tools in 67-272: Application Design and Development

- Anita Borg Institute Faculty Scholarship to attend The 2015 edition of the Grace Hopper Conference
- Best paper award in the 6th International Joint Conference on Natural Language Processing (IJCNLP 2013)
- Full PhD scholarship from the Ministry of Higher Education and Research, France

Service

Conference/Workshop Program Committees

- (2025) Co-organizer of The First Arabic NLP Winter School Co-located with COLING 2025
- (2024-)Associate Editor of the Frontiers in AI Language and Computation Journal.
- (2024) Program Chair of the Second Arabic Natural Language Processing Conference (ArabicNLP 2024)
- (2024) Co-Organizer of the Arabic Financial NLP (AraFinNLP) Shared Task
- (2023-2024) Member in the project "Utilizing digital technology for social cohesion, positive messaging and peace by boosting collaboration, exchange and solidarity" funded by the European Union and co-funded by Friedrich Naumann Foundation in partnership with BoÅşaziÅđi and SabancÅđ Universities. A Project by the Hrant Dink Foundation.
- (2021-2024) Co-organizer of all the editions of the NADI Shared task aiming at distinguishing between Arabic dialects.
- (2023) Program Chair of the Empirical Methods in Natural Language Processing Conference. EMNLP is a leading conference in natural language processing and artificial intelligence. Along with the Association for Computational Linguistics (ACL), it is one of the two primary high-impact conferences for natural language processing research.
- (2023) Co-organizer of the Gender Rewriting Shared Task
- (2022) General Chair of The Seventh Arabic Natural Language Processing Workshop (WANLP 2022).
- (2022) Local co-organizer of the Empirical Methods in Natural Language Processing Conference.
- (2022) Student Volunteer Chair of EMNLP 2022.
- (2022) Keynote Speaker at The Sixth Widening Natural Language Processing Workshop (WiNLP 2022)
- Associate Editor of the ACM Transactions on Asian and Low-Resource Language Information Processing

- Journal reviewer: TALIP 2013, TALIP 2014, Computer Speech and Language, Information Processing and Management, Natural Language Engineering
- Conference reviewer: COLING 2012, AICCSA 2013, IJCNLP 2013, EMNLP 2013, EACL 2014, CICLING 2014, ACL 2014, EMNLP 2014, Semeval 2014, ACL 2015, EMNLP 2015, LREC 16, EMNLP 2016, ACL 2017, EMNLP 2017, LREC 2018, IJCAI 2018, IJCAI 2018, IJCAI 2019, ACL 2019, EMNLP 2019, NAACL 2019, EMNLP 2020, ICIS 2020, EMNLP 2021, ARR 2024
- Program committee member: IJCNLP 2013, AICCSA 2013, AICCSA 2014, ANLP 2014, OSACT2 2016, IJCAI-2019 FinNLP, ACL 2019, IJCAI 2019, WANLP 2020, FNP 2020
- Invited speaker at Social MT 2017, Prague, Czech Republic
- Invited speaker at CEC-TAL 2015, Tunisia
- Publication co-chair: EMNLP 2014 Arabic NLP Workshop, ACL 2015 Arabic NLP Workshop
- Machine Translation session chair: IJCNLP 2013

Thesis and project supervision

- (2024) Boushra Bendou. Independent Study
Topic: AI-based fact-checking on numerical claims for Arabic
- (2024) Malika Dikshit. Senior Honors Thesis
Topic: Exploring Gender Bias in Job Descriptions in STEM Fields: Implications for Female Representation in Science, Technology, Engineering and Mathematics
- (2023) Erin Thomas. Senior Honors Thesis, with Gabriela Gongora-Svartzman
Topic: Exploring Perceptions of Social Movements in Social Media during COVID-19
- (2023) Ahmed Issaoui. Independent Study and QSIURP Summer Research
Topic: Hairstyle Recommendation using Face Recognition Techniques
- (Since 2023) Hasan Alkhder. PhD research work, with Nizar Habash
On Arabic and Turkish Machine Translation
- (Since 2020) Bashar Alhafni. PhD research work, with Nizar Habash
On Automatic gender identification and reinflection in Arabic
- (Since 2019) Go Inoue. PhD research work, with Nizar Habash.
Machine Translation of morphologically rich languages
- (2021) Samir Abdaljalil. Senior Honors Thesis.
Topic: Automatic Summarization of Financial Documents
- (2021) Imane Bendou. Senior Honors Thesis.
Topic: Automatic Arabic Translation of English Educational Content Online using Neural Machine Translation: the Case of Khan Academy

- (2021) Andrew Edward and Abdullah Shaar. Independent Study
Topic: Measuring the Impact of Covid-19 on the shipping industry in the Gulf.
- (2021) Uzair Umar and Saleh Sakir. QSIURP funded project.
Topic: Building a Biomedical parallel corpus
- (2020) Mohammad Shahmeer Ahmed. QSIURP funded project.
Topic: Financial Document Causality Detection
- (2019) Teodora Ostojic. Linguistics Masters internship
Topic: A linguistic analysis of Named Entities in German Financial documents
- (2019) Ana Calatayud. NLP Masters internship
Topic: Automatic Named Entity Recognition in German Financial documents using deep learning.
- (2019) Sonia Ben Berghouth and Ouassama Bouftas. Finance Internship
Topic: Annotation of financial prospectuses
- (2018) Abderrahim Ait-Azzi. Masters Research Project
Topic: Neural Text Generation from Structured Data with application to Finance
- (2016) Maher Hasan Khan, Zehin Khairullah, Ameera Tag Junior Project with Selma Limam Mansar and Chadi Aoun
Topic: GE Power: Data analysis and Modeling for Power Resource Optimization in GE
- (2016) Naassih Gopee thesis with Kemal Oflazer, Bhiksha Raj and William Cohen
Topic: Applying Recurrent Neural Network on Arabic Named Entity Recognition
- (2016) Alaa Khader thesis with Kemal Ofalzer and Francisco Guzmàn
Topic: Computer Assisted Learning for Arabic Speaking ESL Students
- (2015) Sabih Bin Wasi thesis with Kemal Oflazer and Alexander Cheek
Topic: Machine Learning for Twitter Sentiment Analysis
- (2011) Engineer Internship of Jingjing Li with Aurélien Max
Topic: Development of a Web-based interface for writing assistance
- (2010) Msc Internship of Camille Dutrey with Aurélien Max and Delphine Bernhard
Topic: Identifying local modifications and paraphrases in Wikipedia's revision history
- (2010) Engineering Internship of Léton Attanon with Aurélien Max
Topic: Development of a Web-based Game for sub-sentential paraphrase acquisition

International Contributions

- Member of the project "Utilizing digital technology for social cohesion, positive messaging , and peace by boosting collaboration, exchange, and solidarity" funded by the European Union and co-funded by Friedrich Naumann Foundation in partnership with Bogazici and Sabancı Universities. A Project by the Hrant Dink Foundation.
- Nominated by the SIGDAT executive board, the organization that oversees EMNLP (A top-ranked conference in the field of NLP), to run for the Secretary-Treasurer position.

Workshop and Conference Organization

- Program Chair of PoliticalNLP 2024, the Second Workshop on Natural Language Processing for Political Sciences
- Program Chair of ArabicNLP 2024, The Second Arabic Natural Language Processing Conference
- Program Chair of the 2023 Conference on Empirical Methods in Natural Language Processing (EMNLP 2023)
- Part of the Organizing Committee of the 2022 Conference on Empirical Methods in Natural Language Processing (EMNLP 2022)
- Local Organizer of The 2022 Conference on Empirical Methods in Natural Language Processing (EMNLP 2022)
- General Chair of the Seventh Arabic Natural Language Processing Workshop (WANLP 2022)
- Program Chair of PoliticalNLP 2022, the First Workshop on Natural Language Processing for Political Sciences
- Co-organizer of the 3rd Financial Narrative Processing Workshop (FNP 2021)
- Co-organizer of the Sixth Arabic Natural Language Processing Workshop (WANLP 2021)
- Co-organizer of the Fifth Arabic Natural Language Processing Workshop (WANLP 2020)
- Co-organizer of the first Joint Workshop on Financial Narrative Processing and MultiLing Financial Summarisation (FNP-FNS 2020)
- Co-organizer of the 2020 NADI Shared Task on Nuanced Arabic Dialect Identification
- Co-organizer of the FinSBD-2019 Shared Task on Sentence Boundary Detection in PDF Noisy Text in the Financial Domain
- Co-organizer of the MADAR Shared Task on Arabic Fine-Grained Dialect Identification 2019
- Co-organizer of the Fourth Arabic Natural Language Processing Workshop (WANLP 2019)
- Co-organizer of the Third Arabic Natural Language Processing Workshop 2017 (WANLP 2017)
- Co-organizer of the Workshop on Semitic Machine Translation 2016
- Co-organizer of the First Workshop on Arabic Dialect Technologies 2016
- Publication co-chair of the Arabic Natural Language Processing Workshop 2015
- Co-chair of the Second Shared Task on Automatic Arabic Error Correction 2015
- Publication co-chair of the Arabic Natural Language Processing Workshop 2014

University Services

- Member of the use of GenAI in IS education committee (2023-2024)
- Chair of the Meeting of the Minds Committee (2022-2024)
- Member of the CMU-Q Research Council (2022-2024)
- Representative of CMUQ in the Teaching Track Committee, a university-wide the committee looking at Teaching Track issues. (2023)
- Member of the Institutional Review Board (IRB), Carnegie Mellon University in Qatar (Since 2020)
- Member of the Undergraduate Research Committee (2020-2021)
- Member of the Video Policy Committee (2020)
- Advisor of the 2019 CarnegieApps Hackathon
- Member of the Institutional Review Board (IRB), Carnegie Mellon University in Qatar (2016, 2021)
- Member of the Events and Outreach Committee, Carnegie Mellon University in Qatar
- Chaperone of the Silicon Valley Student Innovation Trip, Spring 2016
- Chaperone of the Vietnam Service Learning Trip - May 2017

References

Available upon request