Miguel Ballesteros. PhD. Curriculum Vitae

AI Architect of GenAI at Oracle (OCI)

Computer scientist and computational linguist specializing in the intersection of natural language processing (NLP) and machine learning. My research spans various areas of NLP, language models and its applications.

Personal Email: Miguel Ballesteros - [firstname][lastname]1@gmail.com
Homepage: link

POSITIONS

- AI Architect. Oracle (OCI). New York City, New York, U.S.
  February 2025 -

- Principal Applied Scientist. Amazon Web Services. New York, U.S.
  October 2021 - February 2025

- Senior Applied Scientist. Amazon Web Services. New York, U.S.
  July 2019 - October 2021

- Research Staff Member– IBM Research. New York, U.S.
  October 2016 - July 2019

- Visiting/Assistant Professor – Department of Information and Communication Technologies. Universitat Pompeu Fabra. Barcelona. Spain.
  January 2013 - October 2016

- Visiting Postdoctoral researcher and adjunct instructor/lecturer. Language Technologies Institute. Carnegie Mellon University. Pittsburgh, U.S.
  March 2015 – April 2016

- PhD Student. Department of Software Engineering and Artificial Intelligence. Universidad Complutense de Madrid, Spain.
  September 2010 - December 2012.

EDUCATION

- PhD in Computer Science, Universidad Complutense de Madrid, Spain.
  December 2012.

- Master of Computer Science. Research in Computer Science, NLP Project, Universidad Complutense de Madrid, Spain.
  June 2010. *This may be considered as the first year of the PhD studies.*

- Computer Science. Universidad Complutense de Madrid, Spain.
  June 2009. *This is a 5 year degree (350 ECTS). In the American system it would be equivalent to Bachelor and a Master with a final project.*

PUBLICATIONS, PATENTS, PRESS RELEASES AND AWARDS

PUBLICATIONS

- Over 70 peer-reviewed papers at ACL, EMNLP, NAACL, CL, TACL, EACL and related conferences, cited over 12500 times.

Full list available on Google Scholar and Semantic Scholar.

PATENTS AND PATENTS APPLICATIONS

- US Patent 12086548. Event Extraction from documents with co-reference.

- US Patent 11847406. Natural Language Processing on Semi-Structured Data

- US Patent 11334803. Data processing system to detect neurodevelopmental-specific learning disorders.

- US Patent App. US20220100772A1. Context-Sensitive Linking of Entities to Private Databases.

- US Patent App. 20240202466A1. Adapting Prompts Selected from prompt task collections.

- International Patent App. WO2022072237A1. Lifecycle management for customized natural language processing

PRESS RELEASES

- Amazon Science: link. 2023

- Amazon Science: link. 2022

- MIT News: link. 2019

- Venturebeat.com: link. 2019

- El Periodico. Spanish Newspaper: link. 2017

- ABC. Spanish Newspaper: link. 2016

- El Pais. Spanish Newspaper: link. 2015

AWARDS

- Best Reviewer Award. Selected as one of the Best Reviewers
  Conference on Empirical Methods in Natural Language Processing. 2018.

- Outstanding paper award.
  What do Recurrent Neural Network Grammars Learn about Syntax?
  European Chapter of the Association for Computational Linguistics. April 2017

- Best paper award.
  Towards the understanding of gaming audiences by modeling Twitch emotes
  Third Workshop on Noisy User-generated Text (W-NUT 2017 at EMNLP 2017). September 2017

- Best Student Paper Award
  JADAWeb: A CBR System for Cooking Recipes. Computer Cooking Contest Workshop at the 18th
  International Conference on Case-based Reasoning. **Best Student Paper Award**. (CCC-ICCBR 2010).

ACADEMIC SERVICES

- Action Editor of TACL Journal, 2019 - present

- Action Editor of Computational Linguistics Journal. 2020 - present

- Area Chair. ACL Rolling Review. 2023 - present

- Senior Area Chair, NAACL 2022.

- Tutorial Chair, NAACL 2022.

- Area Chair, EMNLP 2019, CoLING 2018, ACL 2016.

TEACHING AND ACADEMIC SUPERVISION

GRADUATE SUPERVISION

- PhD Thesis. Supervisor of Laura Perez Mayos' PhD Thesis. Thesis: link. Co-supervisor: Leo Wanner. Universitat Pompeu Fabra, Barcelona, Spain. 2022

- 10+ interns mentored at IBM and AWS.

UNDERGRADUATE SUPERVISION

- Supervisor of 3 final Degree Project of undergraduate students University Pompeu Fabra. Barcelona, Spain. 2013-2015

TEACHING EXPERIENCE

- Carnegie Mellon University, Pittsburgh, USA: co-instructor of the Algorithms for Natural Language Processing Course 11711. Carnegie Mellon University. Co-instructors: Chris Dyer and Robert Frederking. Academic Year: 2015-2016

- Universitat Pompeu Fabra, Barcelona, Spain: Instructor of (1) Discrete Mathematics and Graph Theory, main instructor of (2) Software Engineering for Web Applications, and instructor of (3) Databases. Academic Years: 2014-2015, 2013-2014 and 2012-2013

- Universidad Complutense de Madrid, Madrid, Spain: Teaching assistant in the Software Engineering subject of the Computer Science Degree, and (2) Online courses tutor for the Department of Software Engineering and Artificial Intelligence. Faculty of Computer Science. Academic year: 2011-2012.

INVITED TALKS AND PANELS

- AI Trailblazers Summit - Panel (link). New York, United States. June 2024

- Research Seminar - Guest Lecture at Columbia University, Invited by Kathleen Mckeown. New York, United States. November 2022.

- Invited speaker for the Automatic Summarization for Creative Writing workshop at COLING 2022, Gyeongju, Republic of Korea. October 2022.

- Research Seminar at Spain AI, Spain (Virtual). 20th January 2022.

- Research Seminar at Uppsala University, Sweden (Virtual). September 22nd 2021.

- Research Seminar at University Politecnica de Catalunya, Spain (Virtual). September 5th 2020.

- Research Seminar at Xerox Research Centre, France. Host: Xavier Carreras. September 8th 2016. Topic: Transition-based Natural Language Processing.

- UPF-DTIC Research Seminar (Universitat Pompeu Fabra, Barcelona, Spain). Host: Leo Wanner. May 10th 2016. Topic: Deep Learning for Transition-based Natural Language Processing.

- Research Seminar at IBM Watson. Host: Yaser Al-onaizan. April 5th 2016. Topic: Transition-based Natural Language Processing.

- Research Seminar at IBM Almaden. Host: Laura Chiticariu. March 28th 2016. Topic: Transition-based Natural Language Processing.

- Research Seminar at Google New York. Host: Slav Petrov. March 23rd 2016. Topic: Transition-based Natural Language Processing.

- Research Seminar at Carnegie Mellon University. Host: Alan Black. March 10th 2016. Topic: Transition-based Natural Language Processing.

- Research Seminar at Microsoft Research Redmond. Host: Chris Quirk.
  March 3rd 2016. Topic: Greedy Transition-based Dependency Parsing with Stack-LSTMs.

- Research Seminar at Microsoft Research Redmond. Host: Chris Quirk. October 9th 2015. Topic: Greedy Transition-based Dependency Parsing with Stack-LSTMs.

- Seminar at the University of Washington. Host: Noah Smith. October 8th 2015.
  Topic: Greedy Transition-based Dependency Parsing with Stack-LSTMs.

- CLSP student seminar. Johns Hopkins University. Host: Jason Eisner. September 25th 2015. Topic: Greedy Transition-based Dependency Parsing with Stack-LSTMs.

- Seminars of the IXA Research Group. Basque Country University, Spain. Host: Koldo Gojenola. December 19th 2014. Topic: Going to the Roots of Dependency Parsing.

- Machine Learning Lunch Seminars. Carnegie Mellon University. Host: Noah A. Smith.
  October 6th 2014. Topic: Going to the Roots of Dependency Parsing.

- Seminars of the TALP Research Group. Universitat Politècnica de Catalunya, Spain. Host: Xavier Carreras. April 25th 2014. Topic: Going to the Roots of Dependency Parsing.

- Seminars at the Singapore University of Technology and Design, Singapore. Host: Yue Zhang. July 12th 2013. Topic: Deep-Syntacting Parsing [Parsing into Deeper Syntactic Dependencies].

- Seminars of the Department of Linguistics and Philology, Uppsala University, Sweden. Host: Joakim Nivre. March 30th 2012. Topic: MaltOptimizer: A System for MaltParser Optimization.

- Department of Computer Science, Stockholm University, Kista, Sweden. Host: Hercules Dalianis. October 20th 2011. Topic: Enhancing and Applying Dependency Analysis in Different Topics: Negation, Speculation and Text Simplification.

- Seminars of the Department of Linguistics and Philology, Uppsala University, Sweden. Host: Joakim Nivre. October 14th 2011. Topic: Enhancing and Applying Dependency Analysis in Different Topics.

- UCNLG+Eval - The 4th UCNLG Workshop: Language Generation and Evaluation at EMNLP 2011, Edinburgh, United Kingdom. Host: Anja Belz. July 31st 2011. Talk Title: A Proposal for a Spanish Surface Realization Shared Task. (UCNLG+Eval - EMNLP 2011)

LANGUAGES

- Spanish (native)

- English (professional working proficiency)

- Italian (basic)

MISCELLANEOUS

ACADEMIC COMMITTEES

- External Reviewer of Yiqing Zhang's Master Thesis. University of Melbourne, Australia. November 2018.

- Member of Kepa Bengoetxea's PhD examination committee. [Dependency Parsing for Basque].
  Basque Country University, San Sebastian, Spain. 16th December 2014.

- Reviewer for Spanish National Projects. MICINN-ANEP

PROJECTS AND FUNDING

- MIT-IBM: Co-Pi of Hierarchically structured symbolic architectures with neural control for robust and human-like natural language parsing with Prof. Roger Levy. Exploratory Proposal.

- European Project: MULTISENSOR: Mining and Understanding of multilingual content for intelligent sentiment enriched context and social oriented interpretation. (Program: FP7-ICT-610411. European Union)

- European Project: KRISTINA: A knowledge based information agent with social competence and human interaction capabilities. (Program: H2020-645012-RIA. European Union)

SCHOLARSHIP GRANTS

- University Consortium Collaboration Grant: UNED, UCM, UC3M. Spain. May 2010-September 2010.

- Scholarship grant. Department of Informatics Systems and Computation. Complutense University of Madrid, Spain. Granted by the Spanish Ministry of Education and Science. January 2009 - June 2009.

RESEARCH STAYS

- Carnegie Mellon University, USA (1st October 2014 - 9th November 2014)

- Singapore University of Technology and Design (14th June 2013 - 25th July 2013)

- Uppsala University, Sweden (19th March 2012 - 19th April 2012)

- Uppsala University, Sweden (3rd October 2011 - 23rd December 2011)