

Miguel Ballesteros. PhD

Born February 4, 1985 in Madrid, Spain.

Postdoctoral Researcher - Lecturer at Universitat Pompeu Fabra
Visiting Research Scholar at Carnegie Mellon University

Natural Language Processing Group.
Department of Information and Communication Technologies.
Universitat Pompeu Fabra (UPF).
Roc Boronat, 138 08018 Barcelona, Spain

Language Technologies Institute.
School of Computer Science.
Carnegie Mellon University (CMU).
5000 Forbes Avenue, 15213 Pittsburgh, U.S

Phone: (0034) 935421118
Cell phone: (0034) 666164550
Email: miguel.ballesteros@upf.edu
Homepage: <http://miguelballesteros.com>

Qualifications

PhD in Computational Linguistics (Computer Science), Complutense University of Madrid, Spain. December 2012.

Master of Science (60 ECTS). Research in Informatics, NLP Project, Complutense University of Madrid, Spain. June 2010.

Computer Engineering. 5 years degree (350 ECTS). Complutense University of Madrid, Spain. June 2009.

Other Courses

1. Machine Learning for Natural Language Processing. International Master in Language Technology (Prof. Joakim Nivre) . Uppsala University, Sweden. December 2011.
2. Statistical Methods for Natural Language Processing. International Master in Language Technology (Prof. Joakim Nivre). Uppsala University, Sweden. April 2012.

Languages

Spanish (native)
English (proficient)
Italian (intermediate)
Catalan (good comprehension)

Employment

1. Postdoctoral researcher - Lecturer at TALN research group (www.taln.upf.edu) which is part of the Department of Information and Communication Technologies, at Pompeu Fabra University, in Barcelona. Spain. Since January 2013.
2. Visiting Postdoctoral researcher and adjunct instructor/lecturer at the Language Technologies Institute. Carnegie Mellon University. Pittsburgh, U.S. Since March 2015.
3. Researcher. Hired by the NIL research group (nil.fdi.ucm.es). TIN2009-14659-C03-01 Project. Department of Software Engineering and Artificial Intelligence. Complutense University of Madrid. Spain. September 2010-December 2012. During this period I completed my PhD in computational linguistics.
4. Scholarship grant. University Consortium Collaboration Grant: UNED, UCM, UC3M. Spain. May 2010-September 2010.
5. Scholarship grant. Department of Informatic Systems and Computation. Complutense University of Madrid, Spain. Granted by the Spanish Ministry of Education and Science. January 2009 - June 2009.
6. Educational Scholarship grant. Spanish Ministry of Education and Science. Since 2005 to 2009.

Research Interests

Syntax and parsing.

Dependency parsing.

Transition-based modeling for natural language processing.

Deep learning and neural network models for parsing.

International Research Stays

1. 2011. **Uppsala University**, Sweden. Supervised by Joakim Nivre. Phd Student. 3rd October 2011 - 23rd December 2011.
2. 2012. **Uppsala University**, Sweden. Supervised by Joakim Nivre. Phd Student. 19th March 2012 - 19th April 2012.
3. 2013. **Singapore University of Technology and Design**, Singapore. Postdoc. Supervised by Yue Zhang. 14th June 2013 - 25th July 2013.
4. 2013 and 2014. **Birmingham University**, UK. Supervised by Bernd Bohnet. Postdoc. Two short research visits of 2 weeks.
5. 2014. **Carnegie Mellon University**, Pittsburgh, U.S. Supervised by Noah A. Smith and Chris Dyer. Postdoc. 1st October 2014 - 9th November 2014.
6. 2015-16. **Carnegie Mellon University**, Pittsburgh, U.S. Supervised by Noah A. Smith and Chris Dyer. Postdoc. 22 March 2015 - 30th March 2016.

Publications

Journal Papers

1. **Miguel Ballesteros**, Bernd Bohnet, Simon Mille and Leo Wanner. *Data Driven Deep-Syntactic Dependency Parsing*. 2015. **Natural Language Engineering**. 2013 Impact Factor: 0.463

2. **Miguel Ballesteros** and Joakim Nivre. *MaltOptimizer: Fast and Effective Parser Optimization*. 2014. **Natural Language Engineering**. 2013 Impact Factor: 0.463
3. **Miguel Ballesteros** and Joakim Nivre. *Going to the Roots of Dependency Parsing*. 2013. **Computational Linguistics**. 39:1. ISSN 0891-2017. 2013 Impact Factor: 1.468

Refereed Conference Papers

1. Guillaume Lample, **Miguel Ballesteros**, Kazuya Kawakami, Sandeep Subramanian and Chris Dyer. *Neural Architectures for Named Entity Recognition* In proceedings of the Conference of the North American Chapter of the Association for Computational Linguistics (**NAACL-HLT 2016**).
2. Chris Dyer, Adhiguna Kuncoro, **Miguel Ballesteros** and Noah A. Smith. *Recurrent Neural Network Grammars* In proceedings of the Conference of the North American Chapter of the Association for Computational Linguistics (**NAACL-HLT 2016**).
3. **Miguel Ballesteros**, Yoav Goldberg, Chris Dyer and Noah A. Smith. *Training with Exploration Improves a Greedy Stack-LSTM Parser* ArXiv. 2016.
4. Waleed Ammar, George Mulcaire, **Miguel Ballesteros**, Chris Dyer and Noah A. Smith. *Many Languages, One Parser* ArXiv. 2016. .
5. Bernd Bohnet, **Miguel Ballesteros**, Ryan McDonald and Joakim Nivre. *Static and Dynamic Feature Selection in Morphosyntactic Analyzers* ArXiv. 2016.
6. Luz Rello, **Miguel Ballesteros**, Abdullah Ali, Miquel Serra, Daniela Alarcn Snchez and Jeffrey P. Bigham *Dyctective: Detecting Risk of Dyslexia with a Game* PervasiveHealth 2016.
7. **Miguel Ballesteros**, Chris Dyer and Noah A. Smith. *Improved Transition-Based Parsing by Modeling Characters Instead of Words with LSTMs* In proceedings of the Conference on Empirical Methods for Natural Language Processing (**EMNLP 2015**).
8. **Miguel Ballesteros** and Xavier Carreras. *Transition-Based Spinal Parsing*. In proceedings of the Conference of Computational Natural Language Learning (**CoNLL 2015**).
9. Luz Rello, **Miguel Ballesteros** and Jeff Bigham. *A Spellchecker for Dyslexia*. In proceedings of the 17th International ACM SIGACCESS Conference on Computers and Accessibility (**ASSETS 2015**).
10. Chris Dyer, **Miguel Ballesteros**, Wang Ling, Austin Matthews and Noah A. Smith. *Transition-Based Dependency Parsing with Stack Long Short-Term Memory*. In proceedings of Association for Computational Linguistics (**ACL-IJCNLP 2015**).
11. Luz Rello and **Miguel Ballesteros**. *Detecting Readers with Dyslexia Using Machine Learning with Eye Tracking Measures*. In proceedings of W4A2015 (**W4A2015**).
12. Juan Soler-Company, **Miguel Ballesteros**, Bernd Bohnet, Simon Mille and Leo Wanner. *Visualizing Deep-Syntactic Parser Output*. In Proceedings of the Demonstrations of the North American Chapter of the Association for Computational Linguistics (**NAACL-HLT 2015**).
13. **Miguel Ballesteros**, Bernd Bohnet, Simon Mille and Leo Wanner. *Data-driven sentence generation with non-isomorphic trees*. In Proceedings of the North American Chapter of the Association for Computational Linguistics (**NAACL-HLT 2015**).
14. **Miguel Ballesteros**, Bernd Bohnet, Simon Mille and Leo Wanner. *Deep-Syntactic Parsing*. In Proceedings of the 25th International Conference on Computational Linguistics (**COLING 2014**).
15. **Miguel Ballesteros** and Bernd Bohnet. *Automatic Feature Selection for Agenda-Based Dependency Parsing*. In Proceedings of the 25th International Conference on Computational Linguistics (**COLING 2014**).

16. Isabel Ortiz, **Miguel Ballesteros** and Yue Zhang. *ViZPar: a GUI for ZPar with Manual Feature Selection*. Demonstrations of the 30th Spanish Conference on Natural Language Processing (**SEPLN 2014**)
17. **Miguel Ballesteros**, Simon Mille and Leo Wanner. *Classifiers for data-driven deep sentence generation*. In Proceedings of the International Conference of Natural Language Generation (**INLG 2014**).
18. **Miguel Ballesteros**. *Effective Morphological Feature Selection with MaltOptimizer at the SPMRL 2013 Shared Task*. In Proceedings of the EMNLP 2013 Workshop of Statistical Parsing of Morphologically-Rich Languages (**EMNLP-SPMRL 2013**).
19. **Miguel Ballesteros** and Roberto Carlini. *MaltDiver: A Transition-Based Parser Visualizer*. In Proceedings of the Demonstrations of the 6th International Joint Conference on Natural Language Processing (**IJCNLP 2013**).
20. Muntsa Padró, **Miguel Ballesteros**, Hector Martínez and Bernd Bohnet. *Finding dependency parsing limits over a large Spanish corpus*. In Proceedings of the 6th International Joint Conference on Natural Language Processing (**IJCNLP 2013**).
21. **Miguel Ballesteros**, Simon Mille and Alicia Burga. *Exploring Morphosyntactic Annotation Over a Spanish Corpus for Dependency Parsing*. In proceedings of the 2nd international conference on Dependency Linguistics (**Depling 2013**).
22. **Miguel Ballesteros** and Joakim Nivre. *MaltOptimizer: An Optimization Tool for MaltParser*. In Proceedings of the Demonstrations at the 13th Conference of the European Chapter of the Association for Computational Linguistics (**EACL 2012**).
23. **Miguel Ballesteros** and Joakim Nivre. *MaltOptimizer: A System for MaltParser Optimization*. In Proceedings of the 8th International Conference on Language Resources and Evaluation (**LREC 2012**).
24. Alberto Díaz, **Miguel Ballesteros**, Jorge Carrillo de Albornoz, Laura Plaza. *UCM at TREC-2012: Does negation influence the retrieval of medical reports?* In Proceedings of the Text Retrieval Conference - 2012 Medical Records Track (**TREC 2012**).
25. **Miguel Ballesteros**, Virginia Francisco, Alberto Díaz, Jesús Herrera y Pablo Gervás. *Inferring the Scope of Negations in Biomedical Documents*. In Proceedings of the 13th International Conference on Intelligent Text Processing and Computational Linguistics (**CICLING 2012**).
26. Jorge Carrillo de Albornoz, Laura Plaza, Alberto Díaz and **Miguel Ballesteros**. *UCM-I: A Rule-based Syntactic Approach for Resolving the Scope of Negation*. In Proceedings of the 1st Joint Conference on Lexical and Computational Semantics (***SEM 2012**).
27. **Miguel Ballesteros**, Alberto Díaz, Virginia Francisco, Pablo Gervás, Jorge Carrillo de Albornoz and Laura Plaza. *UCM-2: a Rule-Based Approach to Infer the Scope of Negation via Dependency Parsing*. In Proceedings of the 1st Joint Conference on Lexical and Computational Semantics (***SEM 2012**).
28. Pablo Gervás and **Miguel Ballesteros** *A Proposal for a Spanish Surface Realisation Shared Task*. In Proceedings of the 13th European Workshop on Natural Language Generation (**ENLG 2011**).
29. **Miguel Ballesteros**, Virginia Francisco, Alberto Díaz, Jesús Herrera y Pablo Gervás. *Inferring the Scope of Speculation using Dependency Analysis*. In Proceedings of the International Conference on Knowledge Discovery and Information Retrieval (**KDIR 2011**).
30. **Miguel Ballesteros**, Virginia Francisco, Alberto Díaz, Jesús Herrera y Pablo Gervás. *Inferring the Scope of Negation and Speculation via Dependency Analysis*. In Proceedings of the Demonstrations at the 27th Conference of the Spanish Society for Natural Language Processing (**SEPLN 2011**).

31. **Miguel Ballesteros**, Jesús Herrera, Virginia Francisco and Pablo Gervás. *A Feasibility Study on Low Level Techniques for Improving Parsing Accuracy for Spanish Using MaltParser*. In Proceedings of the 6th Hellenic Conference on Artificial Intelligence (**SETN 2010**).
32. **Miguel Ballesteros**, Jesús Herrera, Virginia Francisco and Pablo Gervás. *Towards an N-Version Dependency Parser*. In Proceedings of the 13th International Conference on Text, Speech and Dialogue (**TSD 2010**).
33. **Miguel Ballesteros**, Susana Bautista and Pablo Gervás. *Text Simplification Using Dependency Parsing for Spanish*. In Proceedings of the International Conference on Knowledge Discovery and Information Retrieval (**KDIR 2010**).
34. **Miguel Ballesteros**, Jesús Herrera, Virginia Francisco and Pablo Gervás. *Giving Shape to an N-Version Dependency Parser*. In Proceedings of the International Conference on Knowledge Discovery and Information Retrieval (**KDIR 2010**).
35. **Miguel Ballesteros**, Raúl Martín and Belén Díaz-Agudo. *JADAWeb: A CBR System for Cooking Recipes*. In Proceedings of the Computer Cooking Contest Workshop at the 18th International Conference on Case-based Reasoning. Best Student Paper Award. (**CCC-ICCBR 2010**).

Other Publications

1. **Miguel Ballesteros**. *Exploring Automatic Feature Selection for Transition-Based Dependency Parsing*. 2013. Journal of the Spanish Society for Natural Language Processing. (sepln.org) SEPLN. 51. ISSN 1135-5948.
2. Behrouz Bokharaeian, Alberto Díaz and **Miguel Ballesteros**. *Extracting Drug-Drug interaction from Text using Negation Features* 2013. Journal of the Spanish Society for Natural Language Processing. (sepln.org) SEPLN. 51. ISSN 1135-5948.
3. **Miguel Ballesteros**, Carlos Gómez-Rodríguez and Joakim Nivre. *Optimizing Planar and 2-Planar Parsers with MaltOptimizer*. 2012. Journal of the Spanish Society for Natural Language Processing. (sepln.org) SEPLN. 49. ISSN 1135-5948.
4. **Miguel Ballesteros**, Jesús Herrera, Virginia Francisco and Pablo Gervás. *Enhancing Dependency Analysis by Combining Specific Dependency Parsers*. 2012. Artificial Intelligence and Hybrid Systems. iConcept Press. ISBN: 978-14775547-3-9.
5. **Miguel Ballesteros**, Jesús Herrera, Virginia Francisco and Pablo Gervás. *Are the existing training corpora unnecessarily large?* 2012. Journal of the Spanish Society for Natural Language Processing. (sepln.org) SEPLN. 48. ISSN 1135-5948.
6. **Miguel Ballesteros**, Jesús Herrera, Virginia Francisco and Pablo Gervás. *Analyzing the CoNLL-X Shared Task from a Sentence Accuracy Perspective*. 2012. Journal of the Spanish Society for Natural Language Processing. (sepln.org) SEPLN. 48. ISSN 1135-5948.
7. **Miguel Ballesteros**, Jesús Herrera, Virginia Francisco and Pablo Gervás. *Improving Parsing Accuracy for Spanish Using MaltParser*. 2010. Journal of the Spanish Society for Natural Language Processing. (sepln.org) SEPLN. 44. ISSN 1135-5948.

Dissertations

1. Miguel Ballesteros. *Analyzing, Enhancing, Optimizing and Applying Dependency Analysis*. - (eprints.ucm.es/20086) - Complutense University of Madrid, Spain. (PhD Thesis).
2. Miguel Ballesteros. *Mejora de la precisión para el análisis de dependencias usando Maltparser para el castellano*. - Improving Parsing Accuracy for Spanish Using MaltParser - (eprints.ucm.es/11311) - Complutense University of Madrid, Spain. (MSc).

3. Miguel Ballesteros, José Antonio Blanes and Samer Nabhan. *TALFi: una herramienta para teoría de autómatas y lenguajes formales*. - TALFi: A Tool for Automata Theory and Formal Languages - (eprints.ucm.es/9771) - Complutense University of Madrid, Spain. (BSc).

Invited Talks and Seminars

1. UCNLG+Eval - The 4th UCNLG Workshop: Language Generation and Evaluation at EMNLP 2011, Edinburgh, United Kingdom. Host: Anja Belz.
31st July 2011.
Talk Title: *A Proposal for a Spanish Surface Realization Shared Task*. (UCNLG+Eval - EMNLP 2011)
2. Seminars of the Department of Linguistics and Philology, Uppsala University, Sweden. Host: Joakim Nivre.
14th October 2011.
Topic: *Enhancing and Applying Dependency Analysis in Different Topics*.
3. Department of Computer Science, Stockholm University, Kista, Sweden. Host: Hercules Dalianis.
20th October 2011.
Topic: *Enhancing and Applying Dependency Analysis in Different Topics: Negation, Speculation and Text Simplification*.
4. Seminars of the Department of Linguistics and Philology, Uppsala University, Sweden. Host: Joakim Nivre.
30th March 2012.
Topic: *MaltOptimizer: A System for MaltParser Optimization*.
5. Seminars at the Singapore University of Technology and Design. Host: Yue Zhang.
12th July 2013.
Topic: *Deep-Syntacting Parsing [Parsing into Deeper Syntactic Dependencies]*.
6. Seminars of the TALP Research Group. Universitat Politècnica de Catalunya. Host: Xavier Carreras.
25th April 2014.
Topic: *Going to the Roots of Dependency Parsing*.
7. Machine Learning Lunch Seminars. Carnegie Mellon University. Host: Noah A. Smith.
6th October 2014.
Topic: *Going to the Roots of Dependency Parsing*.
8. Seminars of the IXA Research Group. Basque Country University. Host: Koldo Gojenola.
19th December 2014.
Topic: *Going to the Roots of Dependency Parsing*.
9. CLSP student seminar. Johns Hopkins University. Host: Jason Eisner.
25th September 2015.
Topic: *Greedy Transition-based Dependency Parsing with Stack-LSTMs*.
10. Seminar at the University of Washington. Host: Noah Smith.
8th October 2015.
Topic: *Greedy Transition-based Dependency Parsing with Stack-LSTMs*.
11. Research Seminar at Microsoft Research. Host: Chris Quirk.
9th October 2015.
Topic: *Greedy Transition-based Dependency Parsing with Stack-LSTMs*.

Scientific Software

1. **LSTM Dependency Parser with Character-based Embeddings** - <https://github.com/clab/lstm-parser/tree/char-based> - We propose a parser that uses stack long short-term memory for learning and character-based representations of words with LSTMs- (2015) - In collaboration with Chris Dyer, and Noah A Smith.
2. **LSTM Dependency Parser** - <https://github.com/clab/lstm-parser> - We propose a parser that uses stack long short-term memory for learning - (2015) - In collaboration with Chris Dyer, Wang Ling, Austin Matthews, Noah A Smith.
3. **Deep-Syntactic Generator** - <https://github.com/talnssoftware/deepgenerator/wiki> - We propose a generator that starts from deep (mtt's) syntactic structures as input and generate sentences - (2015) - In collaboration with Bernd Bohnet, Simon Mille and Leo Wanner.
4. **Deep-Syntactic Parser** - <https://github.com/talnssoftware/deepsyntacticparsing/wiki> - We propose a parser that delivers deep (mtt's) syntactic structures as output - (2014) - In collaboration with Bernd Bohnet, Simon Mille and Leo Wanner.
5. **MaltOptimizer** - <http://nil.fdi.ucm.es/maltoptimizer> - A freely available tool developed to facilitate parser optimization using the open-source system MaltParser, a data-driven parser-generator that can be used to train dependency parsers given treebank data - (2012) - In collaboration with Joakim Nivre.
6. **MaltDiver** - <http://taln.upf.edu/pages/MaltDiver/> - MaltDiver is a tool developed to visualize the transitions performed by the transition-based parsers included in MaltParser and to show how the parsers interact with the sentences and the data structures within - (2013) - In collaboration with Roberto Carlini.

Academic Services

Senior Academic Services

Area chair of Tagging, Chunking, Syntax and Parsing of the ACL 2016 Conference. Member of the Program Committee. (<http://acl2016.org/>)

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

Reviewer for Spanish National Projects. MICINN-ANEP research projects (<http://www.mineco.gob.es/portal/site/mineco/>).

Other Academic Services

Reviewer in the syntax, parsing and tagging track for NAACL 2016, ACL 2015, NAACL 2015, ACL 2014, COLING 2014.

Reviewer in the natural language learning track for EMNLP 2015.

Reviewer of LREC 2016, SPMRL 2015, IWPT 2015, DEPLING 2015, LREC 2014, Nodalida 2013, MICAI 2013, CiCling'12, *SEM 2012.

Reviewer of LREV. Language Resources and Evaluation journal.

Reviewer of JAIR. Journal of Artificial Intelligence Research.

Reviewer of NLE. Natural Language Engineering Journal.

Reviewer of CAI. Computing and Informatics Journal.

Member of the organization committee of the SEPLN 2013 - 29th Conference of the Spanish society for natural language processing.

Member of the organization committee and the program committee of the SEPLN 2014 - 30th Conference of the Spanish society for natural language processing.

Member of ACL, SEPLN, SIGNLL, SIGPARSE, SIGDAT.

Teaching

1. Co-instructor of the Algorithms for Natural Language Processing Course 11711. Carnegie Mellon University. With Chris Dyer and Robert Frederking. Academic Year: **2015-2016**
2. Instructor of (1) Discrete Mathematics and Graph Theory and main instructor of (2) Software Engineering for Web Applications. Computer Science Degree for the Department of Information and Communication Technologies, at Pompeu Fabra University, in Barcelona, Spain. Highly evaluated by the students in most of the tasks. Academic Year: **2014-2015**
3. Instructor of (1) Discrete Mathematics and Graph Theory, main instructor of (2) Software Engineering for Web Applications, and instructor of (3) Databases. Computer Science Degree for the Department of Information and Communication Technologies, at Pompeu Fabra University, in Barcelona, Spain. Highly evaluated by the students in most of the tasks. Academic Year: **2013-2014**
4. Instructor of (1) Discrete Mathematics and Graph Theory, main instructor of (2) Software Engineering for Web Applications, and instructor of (3) Databases. Computer Science Degree for the Department of Information and Communication Technologies, at Pompeu Fabra University, in Barcelona, Spain. Highly evaluated by the students in most of the tasks. Academic Year: **2012-2013**
5. Teaching assistant in the Software Engineering subject (Ingeniería del Software) of the Computer Science Degree. Faculty of Informatics. Complutense University of Madrid, Spain. Academic year: **2011-2012**.
6. Online courses tutor for the Department of Software Engineering and Artificial Intelligence. Complutense University of Madrid, Spain. Academic year: **2011-2012**.

Supervision

1. Supervisor of the Final Degree Project of Ivan Latorre (co-supervised with Leo Wanner). Machine Translation through Deep-Syntactic Parsing. Student of the Computer Science Degree. Pompeu Fabra University, in Barcelona, Spain. Defense: July 2015.
2. Supervisor of the Final Degree Project of Isabel Ortiz. ViZPar: a GUI for ZPar with Manual Feature Selection. Student of the Computer Science Degree. Pompeu Fabra University, in Barcelona, Spain. Defense: July 2014.
3. Supervisor of the Final Degree Project of Oscar García. Student of the Computer Science Degree. Pompeu Fabra University, in Barcelona, Spain. Defense: July 2013.