

Wang Ling

CONTACT INFORMATION	Language Technologies Institute Carnegie Mellon University 5503 Gates Hillman Complex Pittsburgh PA 15213	+1 412 320 9515 (United States) +351 93 474 00 88 (Portugal) lingwang@cs.cmu.edu www.cs.cmu.edu/~lingwang
RESEARCH INTERESTS	Machine Translation, Machine Learning, Unsupervised Learning, Natural Language Processing, Social Computing.	
CURRENT OCCUPATION	Carnegie Mellon University, Pittsburgh, United States and Instituto Superior Técnico, Lisbon, Portugal Ph.D. in Language Technologies, Dual Carnegie Mellon Portugal Dual Program, August 2010-Present Advisors: Alan Black, Isabel Trancoso and Chris Dyer. Research Topic: Machine Translation. Cumulative GPA: 3.91 (in 4)	
ACADEMIC BACKGROUND	Instituto Superior Técnico, Lisbon, Portugal Master (MSc) in Information Systems and Computer Engineering, November 2009. Advisor: David Manuel Martins do Matos. Research Topic: Natural Language Processing. Dissertation: "Workflow Control for Distributed Natural Language Processing Applications". Final Grade: 17 (in 20) Degree (BSc) in Information Systems and Computer Engineering, June 2007.	
SCHOLARSHIPS	Scholarship: Portuguese Foundation for Science and Technology (FCT) and Information and Communication Technologies Institute (ICTI), Ph.D. Scholarship, 2010-Present <ul style="list-style-type: none">• Received the Scholarship for the 5-year Dual Ph.D. Carnegie Mellon Portugal Program in Language Technologies.	
TEACHING	Instituto Superior Técnico, Lisbon, Portugal <ul style="list-style-type: none">• Teaching assistant for the "Portfólio Pessoal" course, Fall 2010 and Spring 2011.• Teaching assistant for the "Computer Architecture" course, Spring 2009. Carnegie Mellon University, Pittsburgh, USA <ul style="list-style-type: none">• Teaching assistant for the "Machine Translation"(11-731) course, Spring 2013	
HONORS AND AWARDS	2013-2014 Facebook Graduate Fellowship Finalist	

PROFESSIONAL
SERVICE

Conference Reviewing: North American Association of Computer Linguistics (NAACL), Association of Computer Linguistics (ACL), Empirical Methods in Natural Language Processing (EMNLP), Interspeech.

Monitor LXMLS - Lisbon Machine Learning Summer School, Lisbon, Portugal (2011)

- Cooperation in the writing of the laboratory guide.
- Worked as monitor in laboratory sessions during the summer school.

PROFESSIONAL
EXPERIENCE

Part-time Intern: Spoken Systems Labs (L^2F), INESC-ID, Lisbon, Portugal (2007-2008).

- Development of a Web interface for a local distributed system and also correcting and extending its functionalities.

Part-time Intern: Ericsson, Oeiras, Portugal (2006).

PUBLICATIONS

Peer-reviewed Full Conference Papers

24. *Automatic Keyword Extraction on Twitter* - Luis Marujo, Wang Ling, Isabel Trancoso, Chris Dyer, Alan W. Black, Anatole Gershan, David Martins de Matos, Joao P. Neto, and Jaime Carbonell, In The 53th Annual Meeting of the Association for Computational Linguistics (ACL) 2015, Beijing, China, August 2015
23. *Learning Word Representations from Scarce and Noisy Data with Embedding Subspaces* - Ramón Astudillo, Silvio Amir, Wang Ling, Mario Silva and Isabel Trancoso, In The 53th Annual Meeting of the Association for Computational Linguistics (ACL) 2015, Beijing, China, August 2015
22. *Transition-Based Dependency Parsing with Stack Long Short-Term Memory* - Chris Dyer, Miguel Ballesteros, Wang Ling, Austin Matthews, Noah Smith, In The 53th Annual Meeting of the Association for Computational Linguistics (ACL) 2015, Beijing, China, August 2015
21. *INESC-ID: Sentiment Analysis without hand-coded Features or Linguistic Resources using Embedding Subspaces* - Ramon Astudillo, Silvio Amir, Wang Ling, Bruno Martins, Mario Silva, Isabel Trancoso, In 9th International Workshop on Semantic Evaluation (SemEval 2015), Denver, USA, June 2015
20. *INESC-ID: A Regression Model for Large Scale Twitter Sentiment Lexicon Induction* - Ramon Astudillo, Silvio Amir, Wang Ling, Bruno Martins, Mario Silva, Isabel Trancoso, In 9th International Workshop on Semantic Evaluation (SemEval 2015), Denver, USA, June 2015
19. *Two/Too Simple Adaptations of Word2Vec for Syntax Problems* - Wang Ling, Chris Dyer, Alan W Black, Isabel Trancoso, In Proceedings of the 2015 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Denver, USA, June 2015
18. *Crowdsourcing High-Quality Parallel Data Extraction from Twitter*, Wang Ling, Luis Marujo, Chris Dyer, Alan W Black, Isabel Trancoso, In Proceedings of the Ninth Workshop on Statistical Machine Translation, Baltimore, USA, July 2014
17. *Dual Subtitles as Parallel Corpora*, Shikun Zhang, Wang Ling, Chris Dyer, In Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC'14), Reykjavik, Iceland, May 2014

16. *Linguistic Evaluation Of Support Verb Constructions By OpenLogos And Google Translate*, Anabela Barreiro, Johanna Monti, Brigitte Orliac, Susanne Preuss, Kutz Arrieta, Wang Ling, Fernando Batista, Isabel Trancoso, In Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC'14), Reykjavik, Iceland, May 2014
15. *Paraphrasing 4 Microblog Normalization*, Wang Ling, Chris Dyer, Alan W Black, Isabel Trancoso, In Proceedings of the 2013 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning, Seattle, USA, October 2013
14. *The CMU Machine Translation Systems at WMT 2013: Syntax, Synthetic Translation Options, and Pseudo-References*, Waleed Ammar, Victor Chahuneau, Michael Denkowski, Greg Hanneman, Wang Ling, Austin Matthews, Kenton Murray, Nicola Segall, Yulia Tsvetkov, Alon Lavie, Chris Dyer, In Proceedings of the Eighth Workshop on Machine Translation, Sofia, Bulgaria, August 2013
13. *Microblogs as Parallel Corpora*, Wang Ling, Guang Xiang, Chris Dyer, Alan Black, Isabel Trancoso, In The 51th Annual Meeting of the Association for Computational Linguistics (ACL) 2013, Sofia, Bulgaria, August 2013
12. *Improving Relative-Entropy Pruning using Statistical Significance*, Wang Ling, Nadi Tomeh, Guang Xiang, Alan Black, Isabel Trancoso, In Proceedings of the 25th International Conference on Computational Linguistics (Coling 2012), Mumbai, India, December 2012
11. *Recognition of Named-Event Passages in News Articles*, Luís Marujo, Wang Ling, Anatole Gershman, Jaime Carbonell, João P. Neto, David Matos, In Proceedings of the 25th International Conference on Computational Linguistics (Coling 2012), Mumbai, India, December 2012
10. *Detecting Offensive Tweets via Topical Feature Discovery over a Large Scale Twitter Corpus*, Guang Xiang, Bin Fan, Wang Ling, Jason I.Hong and Carolyn P. Rose, In Proceedings of the 21st ACM International Conference on Information and Knowledge Management (CIKM'12), Hawaii, USA, October 2012
9. *Entropy-based Pruning for Phrase-based Machine Translation*, Wang Ling, João Graça, Isabel Trancoso and Alan W Black, In Joint Conference on Empirical Methods in Natural Language Processing (EMNLP), Jeju, South Korea, July 2012
8. *Overview of computer-assisted language learning for European Portuguese at L2F*, Thomas Pellegrini, Wang Ling, André Freire Silva, Isabel Trancoso, Rui Correia, Jorge Baptista and Nuno J. Mamede, In Proceedings of the International Conference on Computer Supported Education, Porto, Portugal, April 2012
7. *Named Entity Translation using Anchor Texts*, Wang Ling, Pável Calado, Bruno Martins, Isabel Trancoso, Alan Black, Luísa Coheur, In International Workshop on Spoken Language Translation (IWSLT) , San Francisco, USA, December 2011
6. *Discriminative Phrase-based Lexicalized Reordering Models using Weighted Reordering Graphs*, Wang Ling, João Graça, David Martins de Matos, Isabel Trancoso and Alan W Black, International Joint Conference on Natural Language Processing (IJCNLP), Chiang Mai, Thailand, November 2011.

5. *An Agent Based Competitive Translation Game for Second Language Learning*, Wang Ling, Isabel Trancoso, Rui Prada, In SLaTE 2011, ISCA, Venice, Italy, August 2011
4. *Reordering Modeling using Weighted Alignment Matrices*, Wang Ling, Tiago Luís, João Graça, Luísa Coheur, Isabel Trancoso, Association for Computational Linguistics (ACL), Portland, Oregon, USA, June 2011.
3. *BP2EP - Adaptation of Brazilian Portuguese texts to European Portuguese*, Luís Marujo, Nuno Grazina, Tiago Luís, Wang Ling, Luísa Coheur, Isabel Trancoso, In Proceedings of the 15th Conference of the European Association for Machine Translation, European Association for Machine Translation, pages 129-136, Leuven, Belgium, May 2011
2. *The INESC-ID Machine Translation System for the IWSLT 2010*, Wang Ling, Tiago Luís, João Graça, Luísa Coheur, Isabel Trancoso, In International Workshop on Spoken Language Translation, December 2010.
1. *Towards a General and Extensible Phrase-Extraction Algorithm*, Wang Ling, Tiago Luís, João Graça, Luísa Coheur, Isabel Trancoso, In International Workshop on Spoken Language Translation (IWSLT), December 2010.

Technical Reports

1. *Natural Language Engineering on a Computational Grid (NLE-GRID) T3 - Multi-Component Application Builder*, Luís Marujo, Wang Ling, David Martins de Matos, Tech. Rep. 33 / 2008 INESC-ID Lisboa, January 2008.

RESEARCH PROJECTS

Researcher (MT task responsible): PT-STAR (Speech Translation Advanced Research to and from Portuguese), Funded by: FCT (Carnegie Mellon — Portugal Research Projects Program) (2009-2012)

RESOURCES

4. *μtopia: Microblog Translated Post Parallel Corpora* - corpora available at <http://www.cs.cmu.edu/~lingwang/microtopia/>
3. *Geppetto: Toolkit for phrase extraction and scoring for phrase based machine translation* - code available at <http://code.google.com/p/geppetto/>
2. *Relative Entropy Pruning toolkit: Toolkit for pruning phrase tables in Moses* - code available with Moses at <https://github.com/moses-smt/mosesdecoder>
1. *TranslationGame - Translation Game for Second Language Learning* - available at <http://www.12f.inesc-id.pt/TranslationGame/>