

# NANYUN PENG

(310) 448 8250 ◊ npeng@isi.edu

Information Sciences Institute ◊ University of Southern California

4676 Admiralty Way ◊ Marina del Rey, CA 90292

## PROFESSIONAL APPOINTMENTS

---

### University of Southern California

- Research Assistant Professor, Computer Science Department July 2018 - present
- Research Lead, Information Science Institute July 2019 - present
- Computer Scientist, Information Science Institute Sep. 2017 - June 2019

## EDUCATION

---

### Johns Hopkins University

Center for Language and Speech Processing

- Ph.D in Computer Science, advisor: Mark Dredze Sep. 2012 - July. 2017
- *Fred Jelinek Fellowship* 2016

### Peking University

- M.S. in Computer Science, advisor: Houfeng Wang Sep. 2009 - Jul. 2012
- *Google Anita Borg Scholarship* 2011
- Visiting student in NTU Computer Science with *MediaTek Fellowship* 2011
- B.A. in Computational Linguistics and Economics (dual degree) Sep. 2005 - Jul. 2009
- *Excellent Bachelor's Thesis "Automatic Core Concept Extraction from WordNet."* 2009

## RESEARCH INTERESTS

---

**Natural Language Processing:** Creative Language Generation, Multilingual NLP, Information Extraction, Representation Learning for NLP, Finite State Methods for NLP, NLP for Social Good

**Machine Learning:** Deep Neural Networks, Probabilistic Graphical Models, Structured Prediction

## PUBLICATIONS

---

### Peer-reviewed Conferences and Journals

#### 2020

29. Sarik Ghazarian, Ralph Weischedel, Aram Galstyan, **Nanyun Peng**. *Predictive Engagement: An Efficient Metric For Automatic Evaluation of Open-Domain Dialogue Systems*. in Proceedings of the AAAI Conference on Artificial Intelligence (AAAI) 2020. (*acceptance rate: 20.6%*).

#### 2019

28. Rujun Han, Qiang Ning, **Nanyun Peng**. *Joint Event and Temporal Relation Extraction with Shared Representations and Structured Prediction*, in Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing (EMNLP) 2019. (*acceptance rate: 23.8%*).
27. Tao Meng, **Nanyun Peng**, Kai-Wei Chang. *Target Language-Aware Constrained Inference for Cross-lingual Dependency Parsing*, in Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing (EMNLP) 2019. (*acceptance rate: 23.8%*).
26. Emily Sheng, Kai-Wei Chang, Prem Natarajan, **Nanyun Peng**. *The Woman Worked as a Babysitter: On Biases in Language Generation*, in Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing (EMNLP) short 2019. (*acceptance rate: 23.8%*).
25. Xiaolei Huang, Jonathan May, **Nanyun Peng**. *What Matters for Cross-Lingual Named Entity Recognition: An Empirical Analysis*, in Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing (EMNLP) short 2019. (*acceptance rate: 23.8%*).

24. James Mullenbach, Jonathan Gordon, **Nanyun Peng**, Jonathan May. *Do Nuclear Submarines have Nuclear Captains? A Challenge Dataset for Commonsense Reasoning over Adjectives and Objects*, in Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing (EMNLP) *short* 2019. (*acceptance rate: 23.8%*).
23. Xiao Huang, Li Dong, Elizabeth Boschee, **Nanyun Peng**. *Learning A Unified Named Entity Tagger From Multiple Partially Annotated Corpora For Efficient Adaptation*, in Proceedings of the SIGNLL Conference on Computational Natural Language Learning (CoNLL) 2019. (*acceptance rate: 22.7%*).
22. Rujun Han\*, I-Hung Hsu\*, Mu Yang, Aram Galstyan, Ralph Weischedel, **Nanyun Peng**. *Deep Structured Neural Network for Event Temporal Relation Extraction*, in Proceedings of the SIGNLL Conference on Computational Natural Language Learning (CoNLL) 2019. (*acceptance rate: 22.7%*).
21. Wasi Ahmad, Zhisong Zhang, Xuezhe Ma, Kai-Wei Chang, **Nanyun Peng**. *Cross-lingual Dependency Parsing with Unlabeled Auxiliary Languages*, in Proceedings of the SIGNLL Conference on Computational Natural Language Learning (CoNLL) 2019. (*acceptance rate: 22.7%*).
20. Ninareh Mehrabi, Fred Morstatter, **Nanyun Peng**, Aram Galstyan. *Debiasing Community Detection: The Importance of Lowly-Connected Nodes*, in Proceedings of the 2019 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM) 2019.
19. Wasi Ahmad, Zhisong Zhang, Xuezhe Ma, Eduard Hovy, Kai-Wei Chang, **Nanyun Peng**. *On Difficulties of Cross-Lingual Transfer with Order Differences: A Case Study on Dependency Parsing*, in Proceedings of the North American Chapter of the Association for Computational Linguistics (NAACL) 2019. (*acceptance rate: 22.6%*).
18. He He\*, **Nanyun Peng**\*, Percy Liang. *Pun Generation with Surprise*, in Proceedings of the North American Chapter of the Association for Computational Linguistics (NAACL) 2019. (*acceptance rate: 22.6%*).
17. Gully Burns, Xiangci Li, **Nanyun Peng**. *Building Deep Learning Models for Evidence Classification from the Open Access Biomedical Literature*, in DATABASE journal 2019
16. Lili Yao\*, **Nanyun Peng**\*, Ralph Weischedel, Kevin Knight, Dongyan Zhao, Rui Yan. *Plan-And-Write: Towards Better Automatic Storytelling*, in Proceedings of the AAI Conference on Artificial Intelligence (AAAI) 2019. (*acceptance rate: 16%*).

## 2018

15. Xuezhe Ma, Zecong Hu, Jingzhou Liu, **Nanyun Peng**, Graham Neubig, Eduard Hovy. *Stack-Pointer Networks for Dependency Parsing*, in Proceedings of the Annual Meeting of the Association for Computational Linguistics (ACL) 2018. (*acceptance rate: 25.3%*).
14. Mingyue Shang, Zhenxin Fu, **Nanyun Peng**, Yansong Feng, Dongyan Zhao, Rui Yan. *Learning to Converse with Noisy Data: Generation with Calibration*, in Proceedings of the 27th International Joint Conference on Artificial Intelligence (IJCAI) 2018. (*acceptance rate: 25%*).
13. Zhenxin Fu, Xiaoye Tan, **Nanyun Peng**, Dongyan Zhao, Rui Yan. *Style Transfer in Text: Exploration and Evaluation*, in Proceedings of the AAI Conference on Artificial Intelligence (AAAI) 2018. (*acceptance rate: 25%*).

## 2017

12. Dingquan Wang, **Nanyun Peng**, Kevin Duh. *A Multi-task Learning Approach to Adapting Bilingual Word Embeddings for Cross-lingual Named Entity Recognition*, in Proceedings of the International Joint Conference on Natural Language Processing (IJCNLP) *short* 2017. (*acceptance rate: 31%*).
11. **Nanyun Peng**, Hoifung Poon, Chris Quirk, Kristina Toutanova, Wen-Tau Yih. *Cross-Sentence N-ary Relation Extraction with Graph LSTM*, the Transactions of the Association for Computational Linguistics (TACL) 2017, *oral presentation at ACL 2017*.

## 2016

10. **Nanyun Peng**, Mark Dredze. *Improving Named Entity Recognition for Chinese Social Media with Word Segmentation Representation Learning*, in Proceedings of the Annual Meeting of the Association for Computational Linguistics (ACL) short 2016. (acceptance rate: 25.4%)

## 2015

9. Tim Finin, Dawn Lawrie, Paul McNamee, James Mayfield, Doug Oard, **Nanyun Peng**, Ning Gao, Yiu-Chang Lin, Joshi MacKin, Tim Dowd. *HLTCOE Participation in TAC KBP 2015: Cold Start TEDL*, in Proceedings of National Institute of Standards and Technology, Text Analysis Conference (NIST-TAC) 2015.
8. **Nanyun Peng**, Ryan Cotterell, Jason Eisner. *Dual Decomposition Inference for Graphical Models over Strings*, in Proceedings of Conference on Empirical Methods in Natural Language Processing (EMNLP) 2015. (acceptance rate: 24%)
7. **Nanyun Peng**, Mark Dredze. *Named Entity Recognition for Chinese Social Media with Jointly Trained Embeddings*, in Proceedings of Conference on Empirical Methods in Natural Language Processing (EMNLP) short 2015. (acceptance rate: 24%)
6. **Nanyun Peng**, Mo Yu, Mark Dredze. *An Empirical Study of Chinese Name Matching and Applications*, in Proceedings of the Annual Meeting of the Association for Computational Linguistics (ACL) short 2015. (acceptance rate: 22.3%)
5. Ryan Cotterell, **Nanyun Peng**, Jason Eisner. *Modeling Word Forms Using Latent Underlying Morphs and Phonology*, in Transactions of the Association for Computational Linguistics (TACL) 2015, oral presentation at ACL 2015.

## 2014

4. **Nanyun Peng**, Yiming Wang, Mark Dredze. *Learning Polylingual Topic Models from Code-Switched Social Media Documents*, in Proceedings of the Annual Meeting of the Association for Computational Linguistics (ACL) short, 2014. (acceptance rate: 26.1%)
3. Ryan Cotterell, **Nanyun Peng**, Jason Eisner. *Stochastic Contextual Edit Distance and Probabilistic FSTs*, in Proceedings of the Annual Meeting of the Association for Computational Linguistics (ACL) short, 2014. (acceptance rate: 26.1%)

## 2012

2. Tsung-Ting Kuo, San-Chuan Hung, Wei-Shih Lin, **Nanyun Peng**, Shou-De Lin, *Exploiting Latent Information to Predict Diffusions of Novel Topics on Social Networks*, in Proceedings of the Annual Meeting of the Association for Computational Linguistics (ACL) short, 2012. (acceptance rate: 21%)
1. **Nanyun Peng**, Houfeng Wang, *Event Detection in Online News Based on Hierarchical Clustering*, in Proceedings of Chinese National Conference on Computational Linguistics (CNCCL), 2011.

## Referred Workshops and Demos

9. Yiming Wang, Tongfei Chen, Hainan Xu, Shuoyang Ding, Hang Lv, Yiwen Shao, **Nanyun Peng**, Lei Xie, Shinji Watanabe, Sanjeev Khudanpur. *Espresso: A Fast End-to-end Neural Speech Recognition Toolkit*, IEEE Automatic Speech Recognition and Understanding Workshop (ASRU) 2019.
8. Sarik Ghazarian, Johnny Tian-Zheng Wei, Aram Galstyan, **Nanyun Peng**. *Better Automatic Evaluation of Open-Domain Dialogue Systems with Contextualized Embeddings*, in Proceedings of the North American Chapter of the Association for Computational Linguistics (NAACL) *Neural-Gen Workshop* 2019.
7. Rujun Han, Mengyue Liang, Bashar Alhafni, **Nanyun Peng**. *Contextualized Word Embeddings Enhanced Event Temporal Relation Extraction for Story Understanding*, in Proceedings of the North American Chapter of the Association for Computational Linguistics (NAACL) *Workshop on Narrative Understanding* 2019.

6. Seraphina Goldfarb-Tarrant, Haining Feng, **Nanyun Peng**. *Plan, Write, and Revise: an Interactive System for Open-Domain Story Generation*, in Proceedings of the North American Chapter of the Association for Computational Linguistics (NAACL) *system demonstration* 2019.
5. **Nanyun Peng**, Marjan Ghazvininejad, Jonathan May, Kevin Knight. *Towards Controllable Story Generation*, in Proceedings of NAACL Storytelling workshop 2018.
4. **Nanyun Peng**, Mark Dredze. *Multi-task Domain Adaptation for Sequence Tagging*, in Proceedings of the 2nd Workshop on Representation Learning for NLP at the Annual Meeting of the Association for Computational Linguistics (Repl4NLP workshop at ACL) 2017.
3. **Nanyun Peng**, Francis Ferraro, Mo Yu, Nicholas Andrews, Jay DeYoung, Max Thomas, Matthew R. Gormley, Travis Wolfe, Craig Harman, Benjamin Van Durme, Mark Dredze. *A Concrete Chinese NLP Pipeline*, system demo, in Proceedings of the North American Chapter of the Association for Computational Linguistics (NAACL) 2015.
2. Ian E.H. Yen, **Nanyun Peng**, Po-Wei Wang, Shou-de Lin, *On Convergence Rate of Concave-Convex Procedure*, in Proceedings of Advances in Neural Information Processing Systems (NIPS) OPT Workshop, 2012.
1. Wan-Yu Lin, **Nanyun Peng**, Chun-Chao Yen, Shou-de Lin, *Online Plagiarized Detection Through Exploring Lexical, Syntax, And Semantic Information*, system demo, in Proceedings of the Annual Meeting of the Association for Computational Linguistics (ACL), 2012.

## GRANTS

---

- *Evidence Extraction Systems for The Molecular Interaction Literature*, NIH R01 Award, 2017-2021, single PI. \$1,057,064.
- *Guaranteeing AI Robustness against Deception (GARD)*, DARPA, 2019-2023, co-PI (Wael AbdAlmageed PI). \$4,905,684.
- *Understanding Group Biases (UGB)*, DARPA, 2018-2019, co-PI (Aram Galstyan PI). \$898,532.
- *Machine Common Sense (MCS)*, DARPA, 2019-2023, Demo Lead. \$9,338,632. My proportion \$523,520.
- *Knowledge-directed Artificial Intelligence Reasoning Over Schemas (KAIROS)*, DARPA, 2019-2024, Event-ontology Lead. \$7,549,999. My proportion \$475,800.
- *Learning with Less Labels (LwLL)*, DARPA, 2019-2022. \$4,684,994. My proportion \$671,694.
- *Better Extraction from Text Towards Enhanced Retrieval (BETTER)*, IARPA, 2019-2022, Key personnel. \$7,303,285. My proportion \$512,120.

## SELECTED AWARDS

---

<b>Outstanding Area Chair</b>	<i>EMNLP 2019</i>
<b>Fred Jelinek Fellowship</b>	<i>2016</i>
· For Johns Hopkins University students with excellent academic achievements and potential to make new advances in the science and technology of language and speech processing.	
<b>MediaTek Fellowship</b>	<i>2011</i>
· For exchange and visiting programs between universities in mainland China and Taiwan, totally 8 awardees in the whole China and Taiwan area in 2011.	
<b>Google Anita Borg Scholarship</b>	<i>2011</i>
<b>Excellent Bachelor Thesis, Peking University</b>	<i>2009</i>
<b>Peking University Undergraduate Scholarships</b>	<i>2005 - 2008</i>

## SELECTED INVITED TALKS

---

- *Open Domain Generation and Automatic Evaluation*  
**Keynote talk** at the EMNLP Workshop on Neural Generation and Translation *Nov 2019*
- *Creative Generation as Inverse Summarization*  
**Keynote talk** at the EMNLP Workshop on New Frontiers in Summarization *Nov 2019*
- *Towards Creative Text Generation*  
**Keynote talk** at the NAACL Narrative Understanding Workshop *June 2019*
- *Building linguistically informed models for low-resource settings.*  
University of Pennsylvania, Computer Science Department *April 2019*  
Stanford University, Computer Science Department *November 2018*
- *Probabilistically modeling surface forms using latent structure.*  
University of California Los Angeles, Linguistics Department *March 2018*  
Northeast Computational Phonology Circle, University of Delaware *November 2015*
- *Jointly Learning Representations for Low Resource Information Extraction.*  
University of California San Diego, Computer Science Department *February 2018*  
Carnegie Mellon University, Language Technology Institute *January 2018*  
University of California Santa Barbara, Computer Science Department *November 2017*  
University of Southern California, Computer Science Department *Feb 2018*  
University of Utah, School of Computing *August 2017*  
University of Virginia, Department of Computer Science *March 2017*  
Rutgers University, Department of Computer Science *March 2017*  
University of California Davis, Department of Computer Science *March 2017*  
University of Maryland Baltimore County, Department of Computer Science *March 2017*  
University of Southern California, Information Sciences Institute *February 2017*  
University of Massachusetts Amherst, College Information and Computer Sciences *February 2017*

## ADVISEES

---

### Ph.D. Students

- Sidi Lu (Fall 2019 - present)
- Johnny Wei (Fall 2019 - present)
- Mingyu Derek Ma (Fall 2019 - present)
- Te-Lin Wu (Transferred from Joseph Lim, Fall 2019 - present)
- Rujun Han (Co-advised w/ Aram Galstyan, Fall 2018 - present)

### Ph.D. Collaborators

- Emily Sheng (Advised by Prem Natarajan)
- Sarik Ghazarian (Advised by Aram Galstyan)
- I-Hung Hsu (Advised by Prem Natarajan)
- Alexander Spangher (Currently under fellowship rotation)
- Peifeng Wang (Currently under fellowship rotation)

### Interns, Masters, and Undergraduate Students

- Tuhin Chakrabarty (Fall 2019 - present)
- Seraphina Goldfarb-Tarrant (Fall 2018 - present, now Ph.D. at U. Edinburgh)
- Prince Wang, Shufan Wang, Zhifeng Hu (Summer 2019 Interns)
- Mu Yang (Spring 2019 - present)

- Xiangci Li (Summer 2018 - present)
- Xiao Huang (Summer 2018 - Summer 2019)
- Sourav Pati, Bashar Alhafni, Mengyue Liang, Sayan Paul, Haining Feng (Spring 2019)
- James Mullenbach, Xiaolei Huang, Xinyu Wang (Summer 2018 interns)
- Dong Li (Fall 2017 - Spring 2018)

## TEACHING EXPERIENCE

---

### Tutorial: Scalable Construction and Reasoning of Massive Knowledge Bases

- North American Chapter of the Association for Computational Linguistics (NAACL) June, 2018

### University of Southern California, Los Angeles, CA, USA

- *Instructor*, Applied Natural Language Processing, fall, 2019
- *Instructor*, Applied Natural Language Processing, fall, 2018

## PROFESSIONAL ACTIVITIES

---

### Organizer

- The 1st Joint Workshop on Narrative Understanding, Storylines, and Events at ACL July, 2020
- The Second workshop on Reasoning for Complex Question Answering at AACL Feb., 2020
- The Second workshop on Deep Learning for Low-resource NLP at EMNLP Nov., 2019
- The Second Southern California Natural Language Processing Symposium Sep., 2019
- The 24th annual meeting of the intelligent interfaces community (IUI) reception March, 2019
- The Annual WeCNLP (West Coast Natural Language Processing) Summit 2018,2019
- North American Computational Linguistics Olympiad (NACLO) 2013,2018
- MASC-SLL: Mid-Atlantic Student Colloquium on Speech, Language and Learning 2013

### Panel/Mentor/Chair

- The 58th Annual Conference of ACL, *handbook chair* 2020
- The 2nd Annual West Coast Natural Language Processing Summit, *diversity chair* 2019
- AACL NLP-Discourse and Summarization, *session chair* January, 2019
- NIPS Human-Computer Question Answering Competition, *panelist* December, 2017
- NAACL Social Media and Computational Social Science Session, *session chair* December, 2017
- WiNLP (2nd workshop for Widening Natural Language Processing), *mentor* July, 2017

### Program Committee/Reviewer:

- **Area Chair/Senior Program Committee:** EMNLP 19, AACL 20, ACL 20
- **Conferences:** ACL (17,18,19), EMNLP (16,18), NAACL (16,19), AACL (15,16,17,18,19), IJCAI (18,19), COLING (18), CoNLL (17), NLPCC (17,18).
- **Journals:** Computational Linguistics (2017), Informatica (2018), PLOS ONE (2018).