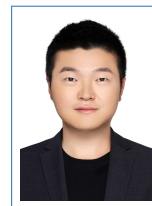


# Siyao (Logan) Peng

Ph.D. candidate in Computational Linguistics  
at Georgetown University

Washington, D.C., USA  
+1-571-519-9051  
✉ [sp1184@georgetown.edu](mailto:sp1184@georgetown.edu)  
in [logan-siyao-peng](https://www.linkedin.com/in/logan-siyao-peng)  
📧 [logan-siyao-peng](mailto:logan-siyao-peng)  
G Siyao Peng  
🔗 [logan-siyao-peng.github.io](https://github.com/logan-siyao-peng)



## Education

- Fall 2017 – present **Ph.D. candidate in Linguistics**, Georgetown University, Washington, DC, USA.  
- Concentration: Computational Linguistics | MS conferred in May 2020  
- Dissertation: Macro-Structural Constraints in RST Dependency Parsing for English and Chinese  
- Advisors: Prof. Amir Zeldes & Prof. Nathan Schneider | Committee member: Prof. Nianwen Xue  
- Defended on December 1, 2022 | Expected Degree Date: Early 2023 | Diploma Date: May 2023
- Fall 2016 – **Ph.D. student in Linguistics**, SUNY - Stony Brook University, Stony Brook, NY, USA.  
Spring 2017 - Advisor: Prof. Thomas Graf
- Fall 2015 – **MA in Linguistics**, Leiden University (Universiteit Leiden), Leiden, the Netherlands.  
Spring 2016 - Concentration: Language Diversity of Africa, Asia and Native America  
- Thesis: Property Concept Words in Six Amazonian Languages | Advisor: Dr. Mily Crevels
- Fall 2011 – **Bachelor of Arts in Linguistics, Applied Mathematics, and French**, University of California - Berkeley, Berkeley, CA, USA.  
Spring 2015

## Papers & Presentations

- Discourse  
Corpus &  
Parsing  
Siyao Peng, Yang Janet Liu, and Amir Zeldes. GCDT: A Chinese RST Treebank for Multigenre and Multilingual Discourse Parsing. In *Proceedings of the 2nd Conference of the Asia-Pacific Chapter of the Association for Computational Linguistics and the 12th International Joint Conference on Natural Language Processing*, pages 382–391, Online only, November 2022. Association for Computational Linguistics [ [paper](#) | [data](#) | [slides](#) ]
- Discourse  
Annotation  
Guideline  
Siyao Peng, Yang Janet Liu, and Amir Zeldes. Chinese Discourse Annotation Reference Manual. Research Report, Georgetown University, October 2022 [ [guideline](#) ]
- Discourse  
Segmentation  
& Relation  
Classification  
Luke Gessler, Shabnam Behzad, Yang Janet Liu, **Siyao Peng**, Yilun Zhu, and Amir Zeldes. DisCoDisCo at the DISRPT2021 Shared Task: A System for Discourse Segmentation, Classification, and Connective Detection. In *Proceedings of the 2nd Shared Task on Discourse Relation Parsing and Treebanking (DISRPT 2021)*, pages 51–62, Punta Cana, Dominican Republic, November 2021. Association for Computational Linguistics [ [paper](#) | [code](#) ]
- Silver Data  
Augmenta-  
tion  
Luke Gessler, **Siyao Peng**, Yang Liu, Yilun Zhu, Shabnam Behzad, and Amir Zeldes. Overview of AMALGUM – Large Silver Quality Annotations across English Genres. *Proceedings of the Society for Computation in Linguistics*, 4(1):434–437, January 2021 [ [abstract](#) | [code](#) ]
- Preposition  
Semantics for  
L2 Speakers  
Michael Kranzlein, Emma Manning, **Siyao Peng**, Shira Wein, Aryaman Arora, and Nathan Schneider. PASTRIE: A Corpus of Prepositions Annotated with Supersense Tags in Reddit International English. In *Proceedings of the 14th Linguistic Annotation Workshop*, pages 105–116, Barcelona, Spain, December 2020. Association for Computational Linguistics [ [paper](#) | [data](#) & [code](#) ]

Machine Translation Quality Estimation	Haijiang Wu, Zixuan Wang, Qingsong Ma, Xinjie Wen, Ruichen Wang, Xiaoli Wang, Yulin Zhang, Zhipeng Yao, and <b>Siyao Peng</b> . Tencent submission for WMT20 quality estimation shared task. In <i>Proceedings of the Fifth Conference on Machine Translation</i> , pages 1062–1067, Online, November 2020. Association for Computational Linguistics [ <a href="#">paper</a>   <a href="#">slides</a> ]
Preposition Semantics in Chinese	<b>Siyao Peng</b> , Yang Liu, Yilun Zhu, Austin Blodgett, Yushi Zhao, and Nathan Schneider. A Corpus of Adpositional Supersenses for Mandarin Chinese. In <i>Proceedings of the 12th Language Resources and Evaluation Conference</i> , pages 5986–5994, Marseille, France, May 2020. European Language Resources Association [ <a href="#">paper</a>   <a href="#">data</a> ]
Silver Data Augmentation	Luke Gessler, <b>Siyao Peng</b> , Yang Liu, Yilun Zhu, Shabnam Behzad, and Amir Zeldes. AMALGUM – A Free, Balanced, Multilayer English Web Corpus. In <i>Proceedings of the 12th Language Resources and Evaluation Conference</i> , pages 5267–5275, Marseille, France, May 2020. European Language Resources Association [ <a href="#">paper</a>   <a href="#">code</a> ]
Dialogue Act Classification	Yue Yu, <b>Siyao Peng</b> , and Grace Hui Yang. Modeling Long-Range Context for Concurrent Dialogue Acts Recognition. <i>Proceedings of the 28th ACM International Conference on Information and Knowledge Management - CIKM '19</i> , pages 2277–2280, 2019. arXiv: 1909.00521 [ <a href="#">paper</a>   <a href="#">poster</a>   <a href="#">slides</a>   <a href="#">code</a> ]
Discourse Segmentation	Yue Yu, Yilun Zhu, Yang Liu, Yan Liu, <b>Siyao Peng</b> , Mackenzie Gong, and Amir Zeldes. GumDrop at the DISRPT2019 Shared Task: A Model Stacking Approach to Discourse Unit Segmentation and Connective Detection. In <i>Proceedings of the Workshop on Discourse Relation Parsing and Treebanking 2019</i> , pages 133–143, Minneapolis, MN, June 2019. Association for Computational Linguistics [ <a href="#">paper</a>   <a href="#">poster</a>   <a href="#">code</a> ]
Preposition Semantics in Chinese	Yilun Zhu, Yang Liu, <b>Siyao Peng</b> , Austin Blodgett, Yushi Zhao, and Nathan Schneider. Adpositional Supersenses for Mandarin Chinese. In <i>Proceedings of the Society for Computation in Linguistics (SCiL 2019)</i> , volume 2, pages 334–337, New York, NY, 2019 [ <a href="#">extended abstract</a>   <a href="#">poster</a> ]
Dependency Conversion	<b>Siyao Peng</b> and Amir Zeldes. All roads lead to UD: Converting Stanford and Penn parses to English Universal Dependencies with multilayer annotations. In <i>Proceedings of the Joint Workshop on Linguistic Annotation, Multiword Expressions and Constructions (LAW-MWE-CxG-2018)</i> , pages 167–177, Santa Fe, NM, 2018 [ <a href="#">paper</a>   <a href="#">poster</a> ]
Multilayer Corpus	<b>Siyao Peng</b> and Amir Zeldes. Validating and Merging a Growing Multilayer Corpus: the Case of GUM. Talk at the 14th American Association for Corpus Linguistics (AACL) Conference, Atlanta, GA, 2018 [ <a href="#">slides</a> ]
Dependency Conversion	<b>Siyao Peng</b> . Should nn be flat or compound: dependency conversion using entity information. Talk at the 2018 Mid-Atlantic Student Colloquium on Speech, Language and Learning (MASC-SLL 2018), Baltimore, MD, 2018
Typology	<b>Siyao Peng</b> . Property Concept Words in Six Amazonian Languages. Master’s thesis, Leiden University, Leiden, the Netherlands, 2016 [ <a href="#">MA thesis</a> ]

---

## Teaching

- Fall 2021 **LING-367: Computational Corpus Linguistics**, *Georgetown University*, Teaching Assistant for Prof. Amir Zeldes.
- Fall 2020 **LING-001: Introduction to Language**, *Georgetown University*, Teaching Assistant for Prof. Paul Portner.
- Spring 2020 **LING-469: Analyzing Language Data with R**, *Georgetown University*, Teaching Assistant for Prof. Amir Zeldes.
- Spring 2020 **LING-462/COSC-482: Statistical Machine Translation**, *Georgetown University*, Teaching Assistant for Dr. Achim Ruopp.
- Fall 2018 **LING-362: Intro: Natural Lang Processing**, *Georgetown University*, Teaching Assistant for Prof. Amir Zeldes.
- Spring 2017 **LIN-230: Languages of the World**, TA, SBU, Teaching Assistant for Prof. Robert Hoberman.
- Fall 2016 **LIN-200: Language in the U.S.**, *Stony Brook University*, Teaching Assistant for Dr. Andrei Antonenko.
- Summer 2013 **MATH-32: Pre-Calculus**, *University of California – Berkeley*, Teaching Assistant for Prof. Richard Gibson.

---

## Reviewing

- Oct 2022 **ACL Rolling Review**, *1 papers reviewed.*
- Aug 2022 **The 2022 Conference on Empirical Methods in Natural Language Processing (EMNLP 2022)**, *5 papers reviewed.*
- July 2021 **Widening Natural Language Processing (WiNLP 2022)**, *3 abstracts reviewed.*
- Jan 2022 **ACL Rolling Review**, *4 papers reviewed.*
- Oct 2021 **Transactions of the Association for Computational Linguistics**, *1 paper co-reviewed with Dr. Nathan Schneider.*
- Sept 2021 **Widening Natural Language Processing (WiNLP 2021)**, *3 papers reviewed.*
- July 2021 **The 2021 Conference on Empirical Methods in Natural Language Processing (EMNLP 2021)**, *4 papers reviewed.*
- May 2021 **The ACL-IJCNLP 2021 Student Research Workshop (ACL-IJCNLP SRW 2021)**, *2 papers reviewed.*
- Mar 2021 **The Joint Conference of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (ACL-IJCNLP 2021)**, *1 paper reviewed.*
- Mar 2021 **The NAACL Student Research Workshop (SRW) 2021**, *1 paper reviewed.*
- Jan 2021 **2021 Annual Conference of the North American Chapter of the Association for Computational Linguistics (NAACL-HLT 2021)**, *2 papers reviewed.*
- Nov 2020 **The 16th Conference of the European Chapter of the Association for Computational Linguistics (EACL 2021)**, *2 papers reviewed.*
- May 2020 **Widening Natural Language Processing (WiNLP 2020)**, *3 papers reviewed.*
- Jan 2020 **The 58th Annual Meeting of the Association for Computational Linguistics (ACL 2020)**, *1 paper reviewed.*
- May 2019 **Widening Natural Language Processing (WiNLP 2019)**, *2 papers reviewed.*

---

## Professional Experience

- Jan 2021 – **NLP Intern**, *PaddlePaddle – Deep Learning Group, Baidu*, Beijing, China.  
June 2021
- May 2020 – **Applied Intern**, *PCG Group – AI Data Center, Tencent*, Beijing, China.  
Sept 2020
- June 2019 – **Educational NLP Intern**, *Educational Applications Group, Pearson*, Boulder, CO, USA.  
Aug 2019

---

## Fellowships and Awards

- Fall 2022 **Linguistics Graduate Student Dissertation Scholarship**, Georgetown University.
- Fall 2017 – **Graduate School Ph.D. Assistantship Stipend**, Georgetown University.  
Spring 2022
- Fall 2017 – **Graduate School Ph.D. Assistantship Stipend**, Georgetown University.  
Spring 2022
- Fall 2019 **Graduate School Student Travel Grants**, Georgetown University.  
Spring 2019
- Fall 2018 **Linguistics Student Travel Grants**, Georgetown University.  
Spring 2019  
Fall 2019  
Fall 2021
- Fall 2016 – **Graduate School Ph.D. Assistantship Stipend**, Stony Brook University.  
Spring 2017
- Summer 2015 **Undergraduate Research Apprentice Program (URAP) Summer Award**, University of California – Berkeley.

---

## Skills

- Programs** Python, R, MATLAB, SQL
- Tools** Linux, Bash, Google Cloud, AWS, Github, L<sup>A</sup>T<sub>E</sub>X
- ML** NumPy, pandas, sklearn, SciPy, Pytorch, Keras, PaddlePaddle, Flair, TensorFlow
- NLP** NLTK, StanfordNLP, spaCy, transformers, PaddleNLP
- Languages** Chinese (native), English (professional), French (intermediate), German (beginner)

---

## Extracurricular Experience

- Sept 2018 – **Georgetown University Chinese Students & Scholars Associations (GUCSSA)**,  
May 2019 *Secretary*, Georgetown University.
- Aug 2012 – **Student Learning Center**, *Math Study Group Leader*, Student Learning Center at  
Dec 2014 University of California – Berkeley.
- Sept 2011 – **Berkeley Chinese Students & Scholars Associations**, *Secretary*, University of California – Berkeley.  
Aug 2012