# Siyao (Logan) Peng

*Postdoctoral Researcher at Munich AI & NLP lab*
*@ Ludwig Maximilian University (LMU) of Munich*

Munich, Bavaria, Germany
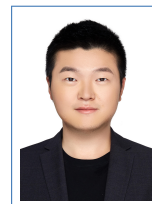+49-160-1433-593
✉ siyao.peng@cis.lmu.de
in logan-siyao-peng
⊕ logan-siyao-peng
G Siyao Peng
⚲ logan-siyao-peng.github.io

## Professional Experience

| | |
|---|---|
| Apr 2023 – Now | **Postdoctoral Researcher**, *Munich AI & NLP lab – LMU Munich*, Munich, Germany.<br>- Research: computational and corpus linguistics, dialect variation, climate change, social media.<br>- Supervisor: Prof. Dr. Barbara Plank |
| Jan 2021 – June 2021 | **NLP Intern**, *PaddlePaddle – Deep Learning Group, Baidu*, Beijing, China. |
| May 2020 – Sept 2020 | **Applied Intern**, *PCG Group – AI Data Center, Tencent*, Beijing, China. |
| June 2019 – Aug 2019 | **Educational NLP Intern**, *Educational Applications Group, Pearson*, Boulder, CO, USA. |

## Education

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

## Papers & Presentations

| | |
|---|---|
| Domain Adaptation Dataset | Tatsuya Aoyama, Shabnam Behzad, Luke Gessler, Lauren Levine, Jessica Lin, Yang Janet Liu, Siyao Peng, Yilun Zhu, and Amir Zeldes. GENTLE: A genre-diverse multilayer challenge set for English NLP and linguistic evaluation. In *Proceedings of the 17th Linguistic Annotation Workshop (LAW-XVII)*, pages 166–178, Toronto, Canada, July 2023. Association for Computational Linguistics [ paper \| data ] |
| Discourse Corpus & Parsing | Siyao Peng. *Cross-Paragraph Discourse Structure in Rhetorical Structure Theory Parsing and Treebanking for Chinese and English.* PhD thesis, Georgetown University, 2023 [ PhD dissertation \| slides ] |

| | |
|---|---|
| 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 Augmentation | 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 **Siyao Peng**. Tencent submission for WMT20 quality estimation shared task. In *Proceedings of the Fifth Conference on Machine Translation*, pages 1062–1067, Online, November 2020. Association for Computational Linguistics [ paper | slides ] |
| Preposition Semantics in Chinese | **Siyao Peng**, Yang Liu, Yilun Zhu, Austin Blodgett, Yushi Zhao, and Nathan Schneider. A Corpus of Adpositional Supersenses for Mandarin Chinese. In *Proceedings of the 12th Language Resources and Evaluation Conference*, pages 5986–5994, Marseille, France, May 2020. European Language Resources Association [ paper | data ] |
| Silver Data Augmentation | Luke Gessler, **Siyao Peng**, Yang Liu, Yilun Zhu, Shabnam Behzad, and Amir Zeldes. AMALGUM – A Free, Balanced, Multilayer English Web Corpus. In *Proceedings of the 12th Language Resources and Evaluation Conference*, pages 5267–5275, Marseille, France, May 2020. European Language Resources Association [ paper | code ] |
| Dialogue Act Classification | Yue Yu, **Siyao Peng**, and Grace Hui Yang. Modeling Long-Range Context for Concurrent Dialogue Acts Recognition. *Proceedings of the 28th ACM International Conference on Information and Knowledge Management - CIKM '19*, pages 2277–2280, 2019. arXiv: 1909.00521 [ paper | poster | slides | code ] |

| | |
|---|---|
| Discourse Segmentation | Yue Yu, Yilun Zhu, Yang Liu, Yan Liu, **Siyao Peng**, Mackenzie Gong, and Amir Zeldes. GumDrop at the DISRPT2019 Shared Task: A Model Stacking Approach to Discourse Unit Segmentation and Connective Detection. In *Proceedings of the Workshop on Discourse Relation Parsing and Treebanking 2019*, pages 133–143, Minneapolis, MN, June 2019. Association for Computational Linguistics [ paper | poster | code ] |
| Preposition Semantics in Chinese | Yilun Zhu, Yang Liu, **Siyao Peng**, Austin Blodgett, Yushi Zhao, and Nathan Schneider. Adpositional Supersenses for Mandarin Chinese. In *Proceedings of the Society for Computation in Linguistics (SCiL 2019)*, volume 2, pages 334–337, New York, NY, 2019 [ extended abstract | poster ] |
| Dependency Conversion | **Siyao Peng** and Amir Zeldes. All roads lead to UD: Converting Stanford and Penn parses to English Universal Dependencies with multilayer annotations. In *Proceedings of the Joint Workshop on Linguistic Annotation, Multiword Expressions and Constructions (LAW-MWE-CxG-2018)*, pages 167–177, Santa Fe, NM, 2018 [ paper | poster ] |
| Multilayer Corpus | **Siyao Peng** 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 [ slides ] |
| Dependency Conversion | **Siyao Peng**. 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 | **Siyao Peng**. Property Concept Words in Six Amazonian Languages. Master's thesis, Leiden University, Leiden, the Netherlands, 2016 [ MA thesis ] |

## Teaching

| | |
|---|---|
| Winter 2023 | **P1: Basismodul Computerlinguistik**, *LMU Munich*, Lecturer. |
| Summer 2023 | **WS: Linguistic Annotation Frameworks**, *LMU Munich*, Lecturer. |
| 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 Spring 2019 | **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

Jun 2023    **ACL Rolling Review**, *1 paper reviewed.*

May 2023    **The 17th Linguistic Annotation Workshop (LAW)**, *3 papers reviewed.*

Mar 2023    **The 61st Annual Meeting of the Association for Computational Linguistics**, *1 paper reviewed.*

Feb 2023    **ACL Rolling Review**, *1 paper reviewed.*

Dec 2022    **ACL Rolling Review**, *3 papers reviewed.*

Oct 2022    **ACL Rolling Review**, *1 paper reviewed.*

Aug 2022    **The 2022 Conference on Empirical Methods in Natural Language Processing (EMNLP 2022)**, *5 papers reviewed.*

July 2022    **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.*

## Fellowships and Awards

Fall 2022    **Linguistics Graduate Student Dissertation Scholarship**, Georgetown University.

Fall 2017 – Spring 2022    **Graduate School Ph.D. Assistantship Stipend**, Georgetown University.

Fall 2017 – Spring 2022    **Graduate School Ph.D. Assistantship Stipend**, Georgetown University.

Fall 2019 Spring 2019    **Graduate School Student Travel Grants**, Georgetown University.

| | |
|---|---|
| Fall 2018<br>Spring 2019<br>Fall 2019<br>Fall 2021 | **Linguistics Student Travel Grants**, Georgetown University. |
| Fall 2016 –<br>Spring 2017 | **Graduate School Ph.D. Assistantship Stipend**, Stony Brook University. |
| 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, LaTeX |
| **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

| | |
|---|---|
| May 2023 –<br>Now | **Munich Chinese Students & Scholars Associations**, *Secretary*, Munich. |
| Sept 2018 –<br>May 2019 | **Georgetown University Chinese Students & Scholars Associations (GUCSSA)**, *Secretary*, Georgetown University. |
| Aug 2012 –<br>Dec 2014 | **Student Learning Center**, *Math Study Group Leader*, Student Learning Center at University of California – Berkeley. |
| Sept 2011 -<br>Aug 2012 | **Berkeley Chinese Students & Scholars Associations**, *Secretary*, University of California – Berkeley. |