Siyao (Logan) Peng

Munich, Germany | +49-160-1433-593 | loganpeng1992@gmail.com |

Professional Summary

AI researcher with three years of postdoctoral experience at LMU Munich and a Ph.D. in Computational Linguistics. Combines expertise in ML, NLP, LLM evaluation, and human-centered AI with hands-on experience designing benchmarks and evaluation frameworks. Skilled in leading interdisciplinary projects, recruiting and mentoring junior researchers, and communicating complex AI concepts to non-AI specialists.

Research Experience

LMU Munich — Munich AI & NLP (MaiNLP) Lab Munich, Germany
Postdoctoral Researcher, *PI: Barbara Plank* Apr 2023–Mar 2026

- Developed benchmarks and explanation-based evaluation methods to improve NLP data quality and assess model reliability, distinguishing annotation errors from valid differences in human judgments[1] and leveraging human-written and LLM-generated explanations[2][3]; resulting work received ACL 2024 and EMNLP 2025 SAC Awards.
- Led a $20K project extending research on human judgment variation to seven languages; co-authored the successful grant proposal, planned the budget, recruited and supervised research assistants, and managed project milestones.
- Investigated NLP and LLM challenges in under-resourced language varieties, collaborating with German dialect specialists to adapt Universal Dependencies (UD) and Named Entity Recognition (NER) methods to Bavarian[4][5]; designed a multilingual QA benchmark to assess LLM knowledge gaps in cross-lingual answer retrieval, showing how translating local-language context into higher-resource languages can improve QA performance[6].
- Collaborated with computer vision and communication science researchers on the Climate Memes project, combining NLP, CV, and social science approaches to analyze stance and framing in climate-change memes; coordinated with partner teams for quarterly progress reviews with the funding agency[7].
- Redesigned introductory NLP courses to make math, statistics, and ML accessible to students from diverse backgrounds through lectures, hands-on exercises, and interactive methods that foster active learning and understanding.
- Mentored doctoral researchers and supervised Bachelor's and Master's theses, providing guidance on research design, experimental setup, and scientific writing, resulting in peer-reviewed conference publications.

Georgetown University Washington, DC, USA
Doctoral Researcher, *Advisors: Amir Zeldes and Nathan Schneider* Aug 2017–May 2023

- Independently built GCDT, a 60K+ token Chinese discourse treebank across five genres for my dissertation; adapted English annotation guidelines to Chinese and improved cross-lingual parsing with RoBERTa through joint training, fine-tuning, and translation-based augmentation[8].
- Engineered rule-based mappings from Stanford Dependencies to Universal Dependencies (UD), integrating entity and coreference information to reduce conversion errors from 1.5% to below 0.5%[9]; contributed to successive GUM corpus releases throughout my Ph.D. through annotation, guideline refinement, and cross-layer consistency checks.
- Implemented UD-based linguistic features and integrated them into GumDrop, an ensemble system combining rule-based, traditional ML, and neural models for robust discourse segmentation across languages and datasets[10].
- Investigated multi-turn dialogue understanding, focusing on long-range conversational context and recognizing multiple intents within a single turn to improve intent recognition in technical support conversations[11].

Internships

Baidu — PaddlePaddle Deep Learning Group Beijing, China Jan–Jun 2021

- Reimplemented the pointer-generator network for abstractive text summarization in PaddlePaddle, adapting an existing PyTorch implementation; the resulting code was later merged into the open-source PaddleNLP library.

Tencent — PCG AI Data Center Beijing, China May–Sep 2020

- Worked with two engineering teams at Tencent on data augmentation and model evaluation for the WMT 2020 English–Chinese quality estimation task; led the paper writing and presented the findings at the WMT workshop.
- Supported Tencent's news platform team in improving news categorization and retrieval by iteratively refining requirements and annotation guidelines, using fine-grained entity classification and concept-based similarity search.

Pearson — Educational Applications Group Boulder, CO, USA Jun–Aug 2019

- Designed a claim–evidence annotation schema for student essays through pilot and think-aloud studies; assessed outsourced labeling quality and explored discourse features in random forest models for automated essay scoring.

Education

Georgetown University — Washington, DC, USA
Ph.D. in Linguistics (Computational Linguistics) — Aug 2017–May 2023

Leiden University — Leiden, Netherlands
M.A. in Linguistics (Language Diversity) — Aug 2015–Aug 2016

University of California, Berkeley — Berkeley, CA, USA
B.A. in Applied Mathematics, French, and Linguistics (Triple Major) — Aug 2011–May 2015

Technical Skills & Languages

Programming & Data: Python, SQL, R, MATLAB, NumPy, pandas, SciPy
ML & DL: PyTorch, TensorFlow, Keras, scikit-learn
NLP & LLMs: Hugging Face (Transformers, Datasets), spaCy, NLTK, Stanza, OpenAI API, Gemini API
Development & Cloud: Git, Linux, Bash, Jupyter, AWS, Google Cloud
Languages: Chinese (native), English (fluent), German (B2), French (intermediate)

Selected Publications

[1] Weber-Genzel, L.*, **Peng, S.***, de Marneffe, M.-C., & Plank, B. (2024). VariErr NLI: Separating Annotation Error from Human Label Variation. ACL 2024. SAC Award: Resources and Evaluation.
[2] Hong, P., Chen, B., **Peng, S.**, de Marneffe, M.-C., & Plank, B. (2025). LiTEx: A Linguistic Taxonomy of Explanations for Understanding Within-Label Variation in Natural Language Inference. EMNLP 2025. SAC Highlight Award.
[3] Zuo, L., Plank, B., & **Peng, S.** (2026). EVADE: LLM-Based Explanation Generation and Validation for Error Detection in NLI. Findings of ACL 2026.
[4] **Peng, S.**, Sun, Z., Shan, H., Kolm, M., Blaschke, V., Artemova, E., & Plank, B. (2024). Sebastian, Basti, Wastl?! Recognizing Named Entities in Bavarian Dialectal Data. LREC-COLING 2024.
[5] Blaschke, V., Kovačić, B., **Peng, S.**, Schütze, H., & Plank, B. (2024). MaiBaam: A Multi-Dialectal Bavarian Universal Dependency Treebank. LREC-COLING 2024.
[6] Pei, R.*, **Peng, S.***, Blaschke, V., Litschko, R., & Plank, B. (2026). Information Asymmetry across Language Varieties: A Case Study on Cantonese-Mandarin and Bavarian-German QA. LREC 2026.
[7] Zhou, S.*, **Peng, S.***, Luebke, S. M., Haßler, J., Haim, M., Mohammad, S. M., & Plank, B. (2025). What Media Frames Reveal About Stance: A Dataset and Study about Memes in Climate Change Discourse. Findings of EMNLP 2025.
[8] **Peng, S.**, Liu, Y. J., & Zeldes, A. (2022). GCDT: A Chinese RST Treebank for Multigenre and Multilingual Discourse Parsing. AACL-IJCNLP 2022.
[9] **Peng, S.**, & Zeldes, A. (2018). All Roads Lead to UD: Converting Stanford and Penn Parses to English Universal Dependencies with Multilayer Annotations. LAW-MWE-CxG 2018.
[10] Yu, Y., Zhu, Y., Liu, Y., Liu, Y., **Peng, S.**, Gong, M., & Zeldes, A. (2019). GumDrop at the DISRPT2019 Shared Task: A Model Stacking Approach to Discourse Unit Segmentation and Connective Detection. DISRPT 2019.
[11] Yu, Y.*, **Peng, S.***, & Yang, G. H. (2019). Modeling Long-Range Context for Concurrent Dialogue Acts Recognition. CIKM 2019.