Yilun **Zhu**

Linguistics · Research Intern

yz565@georgetown.edu | yilunzhu.com | y-l-zhu | yl-zhu

Education

Georgetown University — *Washington D.C.*

Ph.D. in Linguistics — *Aug. 2017 - May 2022*

- Advisors: Dr. Amir Zeldes and Dr. Nathan Schneider

Georgetown University — *Washington D.C.*

M.S. in Linguistics — *Aug. 2017 - May 2019*

Nanjing University — *Nanjing, China*

B.A. in English, Linguistics concentration — *Sep. 2013 - Jun. 2017*

Internships

Microsoft — *Beijing, China*

Software Engineer Intern — *Jun. 2019 - Aug. 2019*

- Worked on semantic parsing for qualitative reasoning Question & Answering in a research team.
- Built a multi-task learning model with Pytorch by extracting qualitative entities and generating logical forms. Trained the model on the QuaRel dataset to predict the correct answer in multiple-choice questions and it out-performed the previous score.

China Telecom — *Beijing, China*

NLP Intern — *May 2018 - Jun. 2018*

- Built a sentiment analysis system, which assists clients to know the feedback of each attributes of car models from customers.
- Prompted an automatic web crawler in Python to collect 110,000+ pieces of positive and negative evaluations of 513 types of cars from online comments; Parsed comments by using StanfordNLP and built a terminology dataset of motor vehicles.
- Trained RNN by using TensorFlow, with pre-trained GloVe word embeddings as input, to predict the sentimental polarity for each comment. The model's F1 achieved 86 assisting to filter out car models with high qualities.

Publications & Presentations

(under review) Knowledge-Informed Coreference Resolution

Y. Zhu

(under review) GUMBY – An Extensible, Genre-Balanced, Richly-Annotated Web Corpus of English

L. Gessler, S. Peng, Y. Liu, **Y. Zhu**, S. Behzad, A. Zeldes

(under review) A Corpus of Adpositional Supersenses for Mandarin Chinese

Y. Liu, S. Peng, **Y. Zhu**, A. Blodgett, Y. Zhao, N. Schneider

Developing Dependency Parsing for English Texts Produced by Multilingual Speakers

Y. Zhu, S. Peng

Georgetown University Round Table (GURT), 2020, Washington DC

GumDrop at the DISRPT2019 Shared Task: A Model Stacking Approach to Discourse Unit Segmentation and Connective Detection

Y. Yu, **Y. Zhu**, Y. Liu, Y. Liu, S. Peng, M. Gong, A. Zeldes

Proceedings of the Workshop on Discourse Relation Parsing and Treebanking (DISRPT) at NAACL-HLT, 2019, Minneapolis, MN

Adpositional Supersenses for Mandarin Chinese

Y. Zhu, Y. Liu, S. Peng, A. Blodgett, Y. Zhao, N. Schneider

Proceedings of the Society for Computation in Linguistics (SCiL) at LSA 2019 Annual Meeting, 2019, New York, NY

Extreme Predicative Adjectives in Mandarin Chinese

Y. Zhu

Presented at 8th International Conference on Formal Linguistics (ICFL-8), 2018, Hangzhou, China

Research

Incorporating Linguistic Structures for Paraphrase Identification with Transformers — *Washington D.C.*

Researcher | Supervisor: Dr. Nathan Schneider — *Sep. 2019 - Dec. 2019*

- Replicate a paraphrase generation model by using a multi-encoder transformer for incorporating a semantic representation SLING from Google.
- Parse Abstract Meaning Representation (AMR) and Universal Dependencies with the state-of-the-art parsers on a paraphrase dataset PAWS.
- Incorporate the structural relations and pre-trained BERT into the multi-encoder transformer and the model outperforms the state of the art on the PAWS benchmark.

Knowledge-Informed Coreference Resolution

Washington D.C.

Researcher | Supervisor: Dr. Amir Zeldes

Sep. 2019 - Nov. 2019

- Extract categorical external knowledge from a knowledge database DBpedia and apply it to the coarse-to-fine (c2f) neural coreference model with BERT.
- The model improves the average F1 score by 0.8% on the BERT-base c2f model on the OntoNotes benchmark.

Evaluation on Universal Conceptual Cognitive Annotation (UCCA) and USim for Paraphrase Detection

Washington D.C.

Researcher

Nov. 2018 - Mar. 2019

- Parsed each pair of paraphrases in MSRP corpus (Microsoft Research Paraphrase Corpus) with TUPA parser to generate UCCA structures for analyzing whether UCCA and USim accurately reflect semantic similarities.
- Built UCCA structure via Tree RNN and add pertained word embeddings as input, evaluating whether semantic structure contributes to paraphrase detection.

Externally configurable reference and non-named entity recognizer (xrenner)

Washington D.C.

Research Assistant | Supervisor: Dr. Amir Zeldes

Feb. 2018 - Nov. 2018

- Established benchmark entities (names, gazetteer, etc.) for a rule-based model in the Chinese subsystem.
- Developed a Logistic Regression classifier with model stacking to predict named entities that are unseen in the corpus by blending rule-based and CRF models, increasing the average F1 score for coreference prediction compared with the previous model.

Skills

NLP & ML	TensorFlow, Pytorch, Keras, Scikit-learn, Numpy, Pandas, StanfordNLP, NLTK
Programming	Proficient in Python, familiar with Java & C/C++, Bash
Miscellaneous	Linux, Google Cloud, AWS, Git, SQL, LaTeX
Languages	Mandarin Chinese (native), English (fluent), French (intermediate)

Honors & Awards

2019	**Linguistics Department Conference Travel Grant**, Georgetown University	*Washington D.C.*
2018	**Linguistics Department Conference Travel Grant**, Georgetown University	*Washington D.C.*
2017	**College Graduate Excellence Award**, Nanjing University	*Nanjing, China*
2015	**The Pacesetter Youth Volunteer**, Nanjing University	*Nanjing, China*
2015	**Renmin Scholarship**, Nanjing University	*Nanjing, China*
2014	**Renmin Scholarship**, Nanjing University	*Nanjing, China*

Teaching

LING-362 Intro to Linguistics

Washington D.C.

Teaching Assistant

Spring 2020

LING-362 Intro to NLP

Washington D.C.

Teaching Assistant

Fall 2019

Related Coursework

Georgetown Univ.	**LING-504** Machine learning for linguistics (Spring 2020) **LING/COSC-462** Statistical machine translation (Spring 2020) **LING/COSC-572** Empirical methods in NLP **ANLY-590** Neural nets & deep learning **LING-461** Speech processing **LING-765** Computational discourse modelling **ANLY-550** Data structures & algorithms
Nanjing Univ.	Calculus, Discrete Mathematics, Basics of Programming (C++)
MOOC	**Introduction to Computer Systems**, [certificate – Nanjing University], Grade: A
Coursera	**Graph Search, Shortest Paths, and Data Structures**, [certificate – Stanford University], Grade: 100%
Coursera	**Divide and Conquer, Sorting and Searching, and Randomized Algorithms**, [certificate – Stanford University], Grade: 100%