

**Ph.D. student in Computational Linguistics @ Georgetown University**

## PUBLICATIONS

- Y. Yu, **S. Peng**, and G. Yang. Modeling Long-Range Context for Concurrent Dialogue Acts Recognition. In *Proc. of CIKM2019*
- Y. Yu, Y. Zhu, Y. Liu, Y. Liu, **S. Peng**, M. Gong, and A. Zeldes. GumDrop at the DISRPT2019 Shared Task: A Model Stacking Approach to Discourse Unit Segmentation and Connective Detection. In *Proc. of DISRPT@NAACL2019*
- Y. Zhu, Y. Liu, **S. Peng**, A. Blodgett, Y. Zhao, and N. Schneider. Adpositional Supersenses for Mandarin Chinese. In *Proc. of SCIL@LSA2019*
- **S. Peng** and A. Zeldes. All Roads Lead to UD: Converting Stanford and Penn Parses to English Universal Dependencies with Multilayer Annotations. In *Proc. of LAW-MWE-CxG@COLING2018*

## Educational Applications of ML/NLP Intern

- Designed an argumentation schema for a wide range of low-stakes high school and college essays based on claim vs. evidence distinctions, by conducting three rounds of pilot annotations and think-aloud experiments.
- Extracted tree-ngrams and computed their KNN and tree edit distance features from automatically parsed Rhetorical Structure Theory (RST) trees and incorporated them into a Random Forest to model the organization trait score of higher-education essays, evaluated on mean squared error.
- Recognized the lack of a promising RST parser that parses plain texts without given gold Elementary Discourse Units (EDUs).

## Discourse Relations across Frameworks

- Evaluated the correlation of implicit relations in paralleled discourse corpora, RST Discourse Treebank (RST-DT) and Penn Discourse Treebank (PDTB), to convert between schemas and parse implicit relations.
- Experimented with regression models for discourse unit segmentation by engineering Universal Dependencies (UD) features in order to remedy the weaknesses of neural models on smaller datasets in the ensemble.

## Coreference & Nested Named-Entity Recognition (NNER)


- Restructured the problem of NNER into detecting entity types and heads and then expanding to its syntactic spans using Flair and Bert embeddings, which reduced the problem from multi-label to single-label classification.
- Augmented multilayered annotations for syntactic and discourse relations, including UD, coreference and RST using StanfordNLP, DPLP and xrenner.

## Semantic Annotations in Multilingual Context

- Supervised annotations of Reddits posted by non-native speakers and planned to analyze their L1 effects on the choice of prepositions.
- Evincd the cross-linguistic applicability of adpositional supersenses to Mandarin Chinese by annotating the Chinese translation of a parallel corpus, *The Little Prince*, and implemented quantitative and qualitative comparisons with annotated English corpora.

## Ph.D. student in Computational Linguistics

📅 August 2017 – present

 **Advisors:** Dr. Amir Zeldes & Dr. Nathan Schneider

**▣** Natural Language Processing, Corpus Linguistics, Discourse Modeling, Dialogue Systems, Machine Translation, Speech Processing, Data Structures & Algorithms, Machine Learning.

Ph.D. student in Linguistics

Stony Brook University, Stony Brook, NY

 August 2016 – May 2017

## M.A. in Linguistics: Language Diversity

Leiden University, Leiden, the Netherlands

Aug 2015 – Aug 2016

## B.A. in Applied Math, French & Linguistics

UC Berkeley, Berkeley, CA

 Aug 2011 – May 2015

 Computing with Data, Probability, Calculus, Linear Algebra & Differential Equations, Discrete Mathematics, Abstract & Complex Analysis.

## LANGUAGES

**Fluent:** Mandarin Chinese, English  
**Familiar:** French

## SKILLS

**NLP & ML** Scikit-learn, Numpy, Pandas, NLTK  
StanfordNLP, Flair, Keras, Pytorch

**Programming:** *Fluent:* Python, R, Bash  
*Familiar:* MATLAB, SQL, Haskell

**Misc** Linux, Google Cloud, Git,  $\text{\LaTeX}$

## AWARDS

**Fall 2019**      Travel Grants for NAACL2019

Spring 2019      Travel Grants for LSA2019

**Fall 2018**      Travel Grants for COLING2018

2017-2022 Georgetown Univ. Ph.D. Stipend

2016-2017      Stony Brook Univ. Ph.D. Stipend

## TEACHING ASSISTANT

Spring 2019 | 2020 Analyzing Lang Data with R  
Fall 2018 Intro: Natural Lang Processing

Spring 2017 Languages of the World

Fall 2016 Language in the U.S.