**Wrangling linguistic data with Python**
**Jacqueline Serigos, *George Mason University***

This workshop will introduce you to the programming language Python and walk you through a typical workflow for converting raw text into an annotated linguistic dataset. We will cover various computational tasks, including reading in raw text files, segmenting text into sentences and tokens, and annotating tokens for various levels of metadata. We will explore the types of linguistic annotation available in the NLP package SpaCy, such as part-of-speech, lemma, and syntactic function. After annotating texts, we will cover techniques for searching and filtering data and use regular expressions to look for word patterns. Lastly, we will touch on the challenges and best practices of working with multilingual data. This workshop is designed to be accessible to both those who are new to programming as well as those who have experience programming.