***Visualizing Russian*: Teaching Vocabulary at the Intersection of Frequency, Grammar, and Communication**

Steven Clancy and Paige Lee, Harvard University

Abstract:

*Visualizing Russian* (https://visualizingrussian.fas.harvard.edu) is a suite of tools benefiting language learners, teachers, and researchers and enabling each user-group to access the complex system of the Russian language through visualization methods in order to leverage the powers of compression and expansion of a massive dataset. Tools developed so far analyze texts for relative difficulty with regard to vocabulary content, create frequency lists and display word forms used for each lemma in a text, compare target vocabulary to coverage in a text, identify the most commonly used word forms for a lemma, compare the relative frequency of near-synonyms or other items in a semantic domain, show the usage of nouns with various cases or verbs with various person/number/tense/aspect combinations, and identify the case governance or preposition usage of Russian verbs. Additional tools break down word-formation for prefixes-roots-suffixes in words and identify the field of morphologically-related or semantically-related words for any target word. These tools have been developed in collaboration with an undergraduate researcher studying computer science and Slavic languages. With input and guidance from the rest of the team, she has designed, prototyped, tested, and deployed tools to the *Visualizing Russian* site, in addition to contributing to the maintenance and development of the dataset. She has also analyzed Russian language data sources such as the abridged Russian National Corpus to bring “real world” statistics and examples to the tools to demonstrate actual language usage. In the process of developing these tools, she has explored novel web visualization techniques for language data; gained a deeper knowledge of the intricacies of Russian morphology, semantics, syntax, and grammar; applied concepts in statistical language analysis such as word embeddings and principal component analysis; and learned about the role that digital pedagogical tools can play in Russian language learning.