Current project is based on the pre-collected [dataset](#) from Russian National Corpus. The authors of the dataset are Tore Nesset, Laura A. Janda and Mihail Kopotev. The objective of the dataset and my project is to examine NOM-INS construction in Russian like "дурак дураком" and "вода водой". This dataset consists of two pasts: the first one examines plural forms of such constructions and the second one – singular forms. Both parts have the same set of features: semantics of the construction (abundance, extreme, instrument, paragon, and topic change), syntactic role of the noun (Ins verb-governed, Nom verb-governed, Nom/Ins verb governed and ungoverned), semantic classes of nouns (abstract, concrete, human + animal, substances), pejorative (as involving or not involving a pejorative nuance), chained (marks cases with 2+ occurrences of constructions), associated verb, left and right contexts, construction itself, punctuation marks immediately following the construction, metadata (title of the work, name and birthday of the author, header, year of creation, "sphere" and "type" of the text, topic, publication name and year), full context.

This list of features will be expanded. Primarily, the morphological parameters of the associated verb are going to be added. Also, using context and morphology, I am going to create patterns of the construction's usage.

My goal is to examine what features have the best correlation to the semantics of the construction. The hypothesis is that syntactic role and semantic class of nouns will be most relevant features for the construction classification. Linear regression model and T-test will be used for the hypothesis testing.