

The goal of my research is to explore the behavior of rival forms of construction describing transportation in Russian. There are two possible ways to specify the type of transport used for movement:

- construction with a noun in Instrumental case (e.g. *доехать поездом, доставлять самолетами, путешествовать автобусом* etc.)
- construction with a preposition and a noun in Prepositional case (e.g. *добираться на поезде, лететь на самолете, уезжать на автобусе* etc.)

Thus, the study aims to find out whether the choice of the construction depends on certain specific properties of the context, which it used in, or not.

## Hypothesis

Null hypothesis of the study: There is no association between the choice of the construction and certain contextual variables (such as properties of the verb in the construction, number of the noun, length of sentence and year of creation).

H1: There is an association between the choice of the construction and certain contextual variables.

## Data

For the purpose of the study I collected examples from Russian National Corpus containing constructions mentioned above. I annotated examples in order to identify contextual variables, which supposedly can influence construction choice. The dataset contents following columns:

- Verb – verbal part of the construction
- Transport – noun part of the construction
- Construction\_type – one of two types of the construction
- Normal\_form - normal form of the verb
- Prefix – whether or not the verb has prefix
- Tense – tense of the verb
- Aspect – aspect of the verb
- Number – number of the transport noun
- Full\_context – complete sentence with the construction
- Sent\_length – the length of the sentence
- Author – name of the author
- Header – name of the source of sentence
- Created – year of creation of the source.

## Experiment design

In order to test the hypothesis, I will conduct statistical tests that will determine whether there is an association between variables. The target variable is *Construction\_type* and predictor variables are *Prefix*, *Tense*, *Aspect*, *Number*, *Sent\_length* and *Created*.

I will use the linear regression model to predict *Construction\_type* by the other variables and to estimate the size of the effect of each variable. It's also possible to explore each categorical variable independently using Chi-squared test.