Analysis of the Standard Russian data

This appendix will not be included in the final version.

Preparation of the data

```
library("tidyverse")
```

Import the data in R

```
gen_std <- read.csv("std_rus.csv", stringsAsFactors=TRUE)
```

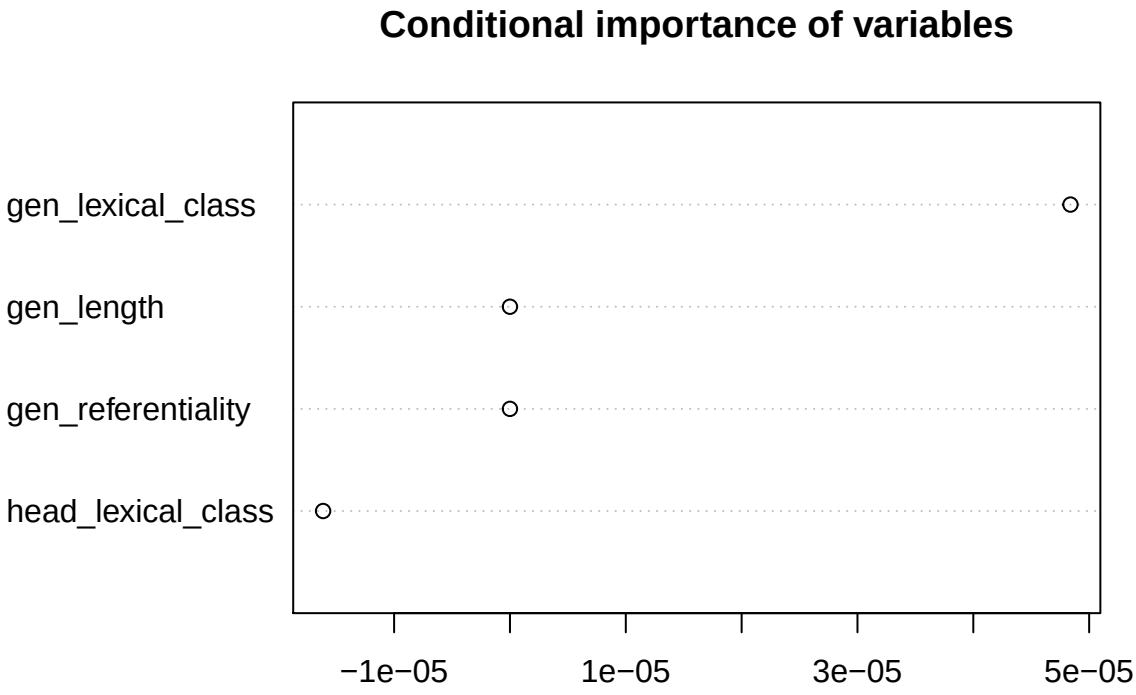
Set correct reference levels

```
gen_std$head_lexical_class <- relevel(gen_std$head_lexical_class, "non_kinship")
gen_std$gen_lexical_class <- relevel(gen_std$gen_lexical_class, "non_human")
gen_std$gen_referentiality <- relevel(gen_std$gen_referentiality,"non_definite")
gen_std$gen_length <- relevel(gen_std$gen_length,"one-word")
```

Random forest

```
library("party")
```

```
gen_std_rf <- cforest(position ~ gen_lexical_class + head_lexical_class +
                        gen_referentiality + gen_length,
                      data = gen_std, controls = cforest_unbiased(ntree = 500, mtry = 2))
gen_std_varimp <- varimp(gen_std_rf, conditional = TRUE)
dotchart(sort(gen_std_varimp), main = "Conditional importance of variables")
```


Conditional importance of variables
gen_lexical_class
gen_length
gen_referentiality
head_lexical_class
−1e−05
1e−05
3e−05
5e−05