R-code used in the paper "A non-standard syntax of numerals in the Russian speech of Nanai and Ulcha speakers"

#Loading the dataset

```
> setwd("D:/???")
> numerals <- read.csv("tung_rus_numerals.csv", header = TRUE)
```

Releveling process (reference levels → those predisposing to the genitive encoding)

```
> numerals$pattern <- relevel(numerals$pattern, ref = "rus")
> numerals$num_type <- relevel(numerals$num_type, ref = "gen_pl")
> numerals$num_semantics <- relevel(numerals$num_semantics, ref = "large")
> numerals$noun_semantics <- relevel(numerals$noun_semantics, ref = "time&measure")
```

#Logistic regression model

Initial full model

#The initial model includes the following predictors: num_type (syntactic type of the numeral: gen_pl, gen_sg), num_semantics (semantics of the numeral: large, basic), and noun_semantics (semantics of the noun: time&measure, non_time&measure)

```
> glm_numerals <- glm (pattern ~ num_type + num_semantics + noun_semantics, data = numerals, family ="binomial")
> summary(glm_numerals)
```

#all predictors except for num_type appear to have a significant effect

```
Call:
glm(formula = pattern ~ num_type + num_semantics + noun_semantics,
    family = "binomial", data = numerals)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-1.0330  -0.9860  -0.4545  -0.2488   2.6422

Coefficients:
                              Estimate Std. Error z value Pr(>|z|)
(Intercept)                    -3.4598     0.3987  -8.677  < 2e-16 ***
num_typegen_sg                  0.1189     0.2793   0.426   0.6704
num_semanticsbasic              1.1229     0.3972   2.827   0.0047 **
noun_semanticsnon_time&measure  1.8683     0.3557   5.252  1.5e-07 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 459.65  on 430  degrees of freedom
Residual deviance: 384.94  on 427  degrees of freedom
AIC: 392.94

Number of Fisher Scoring iterations: 5

Call:
glm(formula = pattern ~ num_type + num_semantics + noun_semantics,
    family = "binomial", data = numerals)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-1.0330  -0.9860  -0.4545  -0.2488   2.6422

Coefficients:
                              Estimate Std. Error z value   Pr(>|z|)
(Intercept)                    -3.4598     0.3987  -8.677   < 2e-16 ***
num_typegen_sg                  0.1189     0.2793   0.426    0.6704
num_semanticsbasic              1.1229     0.3972   2.827    0.0047 **
```

```
noun_semanticsnon_time&measure   1.8683     0.3557   5.252   1.5e-07 ***

---
Signif. codes:
0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 459.65  on 430  degrees of freedom
Residual deviance: 384.94  on 427  degrees of freedom
AIC: 392.94

Number of Fisher Scoring iterations: 5
```

#Tuning the model using the Akaike Information Criterion (AIC)

#The predictors will be dropped from the initial model one by one. The predictor can be dropped (=irrelevant for the model) if its absence reduces the AIC value.

```
> drop1(glm_numerals)
```

#the model without num_type has a smaller AIC value (391.12) than the initial one (392.94)

```
Single term deletions

Model:
pattern ~ num_type + num_semantics + noun_semantics
             Df Deviance    AIC
<none>              384.94 392.94
num_type      1   385.12 391.12
num_semantics   1   393.67 399.67
noun_semantics  1   420.33 426.33
```

#Model 1: includes num_semantics and noun_semantics (num_type was dropped)

```
> glm_numerals1 <- glm (pattern ~ num_semantics + noun_semantics, data = numerals, family ="binomial")
> summary(glm_numerals1)
Call:
glm(formula = pattern ~ num_semantics + noun_semantics, family = "binomial",
    data = numerals)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-1.0111  -1.0111  -0.4485  -0.2510   2.6357

Coefficients:
                            Estimate Std. Error z value Pr(>|z|)
(Intercept)                 -3.4419     0.3953  -8.706  < 2e-16 ***
num_semanticsbasic           1.1956     0.3583   3.337 0.000847 ***
noun_semanticsnon_time&measure   1.8417     0.3503   5.257 1.46e-07 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 459.65  on 430  degrees of freedom
Residual deviance: 385.12  on 428  degrees of freedom
AIC: 391.12

Number of Fisher Scoring iterations: 5
```

#Model 1 is optimal, since no more predictors can be dropped: the model with no dropping has the smallest AIC value (391.12)

```
> drop1(glm_numerals1)
Single term deletions

Model:
pattern ~ num_semantics + noun_semantics
               Df Deviance    AIC
<none>              385.12 391.12
num_semantics   1   398.09 402.09
noun_semantics  1   421.24 425.24
```