

# Проект по R: партитивное использование формы генитива с местоимениями

Данилова В., Пономарева М.

## Теория

### Предмет исследования:

формы генитива и аккузатива неопределенного местоимения “что-то” в DO при переходных глаголах.

Для исследования было выбрано **пять** переходных глаголов(каждой форме несовершенного вида сопоставлена одна форма совершенного):

- ждать/подождать
- хотеть/захотеть
- требовать/потребовать
- искать/поискать
- просить/попросить

### Примеры интересующих нас случаев:

*Сергей все чего-то ждал.*

*Она потребовала от него что-то невозможное.*

*Каждый ищет что-то свое.*

*Попросили что-то простое и незамысловатое.*

### Выбор глаголов:

Чередование форм генитива и аккузатива встречается у целого ряда глаголов с семантикой физического восприятия и воздействия, перемещения, обладания, речевой деятельности, мыслительной и эмоциональной деятельности. Выбранные нами глаголы можно отнести к последней семантической группе. Мы намеренно ограничили выбор глаголами с близкой семантикой, предварительно оценив частотность интересующих нас случаев в корпусе.

### Выбор объекта:

Известна зависимость выбора падежной формы от семантики DO (“хочу яблок/яблоки”, “налить молока/молоко”, на выбор влияет исчисляемость объекта, значение партитивности, определенность и т.д.). Исследовать значимость этих факторов представляется нам интересной задачей, требующей больших усилий в первую очередь связанных с разметкой данных. В нашей работе мы сознательно ограничиваемся неопределенным местоимением в роли DO, повышая таким образом вероятность использования партитивного генитива при глаголе.

## Датасет

Мы проверяли влияние следующих факторов на выбор падежных форм: - вид глагола (совершенный vs несовершенный)

- время глагола (прошедшее vs настоящее/будущее)
- наличие прилагательного при местоимении
- позиция местоимения относительно глагола (постпозиция vs препозиция)
- глагольная пара

Четыре из перечисленных факторов являются бинарными, последний имеет 5 значений. Для всех возможных 80 комбинаций факторов мы искали примеры в НКРЯ, в датасет вошли все найденные релевантные примеры (1882 случая).

```
library(readxl)
data_project <- read_excel("data_project_tf.xlsx")
data_project[1555:1585,]
```

##	Acc	past	perfect	adjective	postposition	root
## 1555	TRUE	TRUE	FALSE	FALSE	FALSE	проси
## 1556	TRUE	TRUE	FALSE	FALSE	FALSE	проси
## 1557	TRUE	TRUE	FALSE	FALSE	FALSE	проси
## 1558	TRUE	TRUE	FALSE	FALSE	FALSE	проси
## 1559	TRUE	TRUE	FALSE	FALSE	FALSE	проси
## 1560	TRUE	TRUE	FALSE	FALSE	FALSE	проси
## 1561	TRUE	TRUE	FALSE	FALSE	FALSE	проси
## 1562	TRUE	TRUE	FALSE	FALSE	FALSE	проси
## 1563	TRUE	TRUE	FALSE	FALSE	FALSE	проси
## 1564	TRUE	TRUE	FALSE	FALSE	FALSE	проси
## 1565	TRUE	TRUE	FALSE	FALSE	FALSE	проси
## 1566	TRUE	TRUE	FALSE	FALSE	FALSE	проси
## 1567	TRUE	TRUE	FALSE	FALSE	FALSE	проси
## 1568	TRUE	TRUE	FALSE	FALSE	FALSE	проси
## 1569	TRUE	TRUE	FALSE	FALSE	FALSE	проси
## 1570	TRUE	TRUE	FALSE	FALSE	FALSE	проси
## 1571	TRUE	TRUE	FALSE	FALSE	FALSE	проси
## 1572	FALSE	TRUE	TRUE	FALSE	TRUE	проси
## 1573	TRUE	TRUE	TRUE	FALSE	FALSE	проси
## 1574	TRUE	TRUE	TRUE	FALSE	FALSE	проси
## 1575	TRUE	TRUE	TRUE	FALSE	TRUE	проси
## 1576	TRUE	TRUE	TRUE	TRUE	TRUE	проси
## 1577	FALSE	FALSE	FALSE	FALSE	FALSE	требова
## 1578	FALSE	FALSE	FALSE	FALSE	FALSE	требова
## 1579	FALSE	FALSE	FALSE	FALSE	FALSE	требова
## 1580	FALSE	FALSE	FALSE	FALSE	FALSE	требова
## 1581	FALSE	FALSE	FALSE	FALSE	FALSE	требова
## 1582	FALSE	FALSE	FALSE	FALSE	FALSE	требова
## 1583	FALSE	FALSE	FALSE	FALSE	FALSE	требова
## 1584	FALSE	FALSE	FALSE	FALSE	FALSE	требова
## 1585	FALSE	FALSE	FALSE	FALSE	FALSE	требова
##						
context						
## 1555						Огурцов пылил сапогами на летучке: что-то просил, что-то тре
бовал, в чём-то						

## 1556 папа, как они краем уха...  
 что-то просил.  
 ## 1557 Все что-то просили у святых, сует  
 ливо крестились  
 ## 1558 болтами на них, как будто что-то просил  
 он у солнца или  
 ## 1559 <U+2015> Ну, если она что-то тебя просила или прик  
 азывала сделать  
 ## 1560 Не могу сказать, что я что-то когда-то  
 просил у него.  
 ## 1561 предыдущей их акции, иногда Нина что-то просила дополнить  
 , а потом Игорь  
 ## 1562 что вы подавали какие-то документы. Что-  
 то просили. Он.  
 ## 1563 делах, приветствовали, жаловались, сообщали новости,  
 что-то просили...  
 ## 1564 бабы с детьми, то ли что-то просили, то ли тор  
 говали, мальчик  
 ## 1565 просила у меня в сентябре. Что-то просила. Что-т  
 о ей было нужно  
 ## 1566 вашем парохде поломка и вы что-то про  
 сили у старика.  
 ## 1567 нет ни одного человека, который что-то просил бы  
 и что-то делал  
 ## 1568 Братья тормошили его, что-то просили, требова  
 ли, но он сидел  
 ## 1569 поговорить с ним наедине. Он что-то там просил у  
 кардинала. Так  
 ## 1570 нему подошел старик, сдернул картузик, <U+2015> что-то ,  
 видимо, просил.  
 ## 1571 Он что-то просил, она что-то  
 делала, встала  
 ## 1572 из тысяч, единственная, в которой я попросил чего-то  
 для себя самого  
 ## 1573 Если я что-то один раз попроси  
 л, завтра будет  
 ## 1574 когда вы что-то попросил  
 и тогда у меня.  
 ## 1575 Там , попросив еще что-  
 то, он, лавируя  
 ## 1576 на него с отвращением, точно он попросил у нее что  
 -то неприличное  
 ## 1577 вы ведёте себя неправильно, вы чего-то требуете, вы  
 капризничаете...  
 ## 1578 наших детей, ругаем, «достаем», поучаем, чего-то от них требуем,  
 ждем, воплощаем  
 ## 1579 мучая захваченное, что-то вымогая и чего-то требуя, то пропад  
 ая, то возникая  
 ## 1580 требует жертв или что любовь чего-  
 то там требует?

```
## 1581 Я же от всех чего-то требую: и что
бы глаза горели
## 1582 честное слово. Каждый день все чего-то требуют. Сдел
ай то, подай то
## 1583 с ее ребенком и неявно чего-то требующая, и требую
щая все сильнее
## 1584 недоделанности, вернуться туда должна, закончить? Что-то зовет, требует. К
акая-то ошибка?
## 1585 Потом звонит по телефону, чего-то требует, приказы
вает, угрожает.
```

## Гипотеза

Наша гипотеза: выбор падежного оформления DO зависит от глагола, факторы вида, времени, прилагательного и позиции местоимения могут в разной степени влиять на разные глагольные пары.

## Графики, отображающие распределение наших данных

```
library(tidyverse)
```

```
## Loading tidyverse: ggplot2
## Loading tidyverse: tibble
## Loading tidyverse: tidyr
## Loading tidyverse: readr
## Loading tidyverse: purrr
## Loading tidyverse: dplyr
```

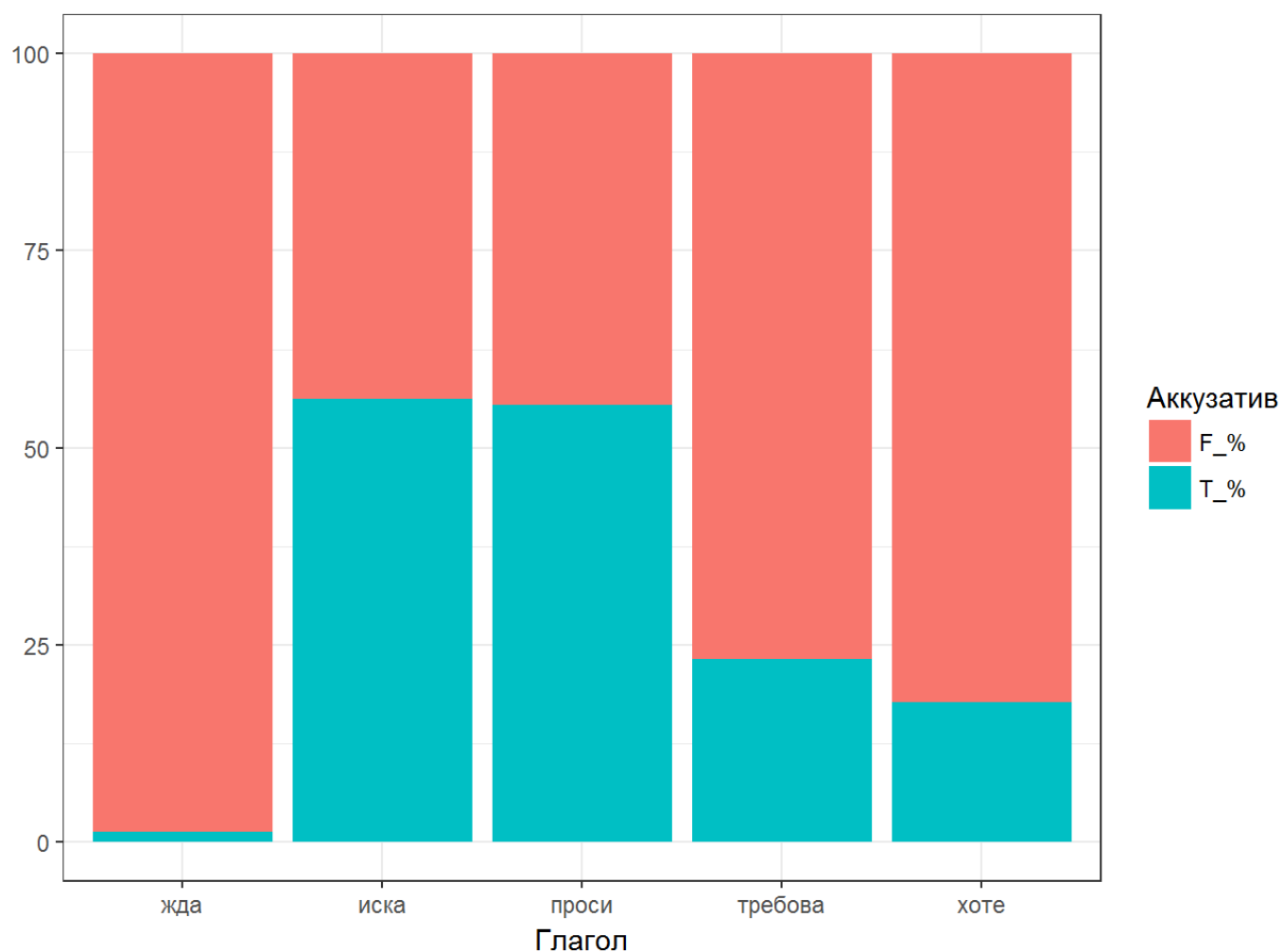
```
## Conflicts with tidy packages -----
```

```
## filter(): dplyr, stats
## lag(): dplyr, stats
```

```
data_project %>%
  select(root, Acc) %>%
  count(root, Acc) %>%
  spread(key = Acc, value = n) %>%
  mutate(sum = `TRUE`+`FALSE`,
         `T_%` = `TRUE`/sum*100,
         `F_%` = `FALSE`/sum*100) %>%
  select(-c(`TRUE`, `FALSE`, sum)) %>%
  gather(key = Acc, value = percent, `T_%`:`F_%`) %>%
  na.omit() -> data_rel
```

Общее соотношение употребления аккузатива и генитива для каждого из корней отображено на следующем графике:

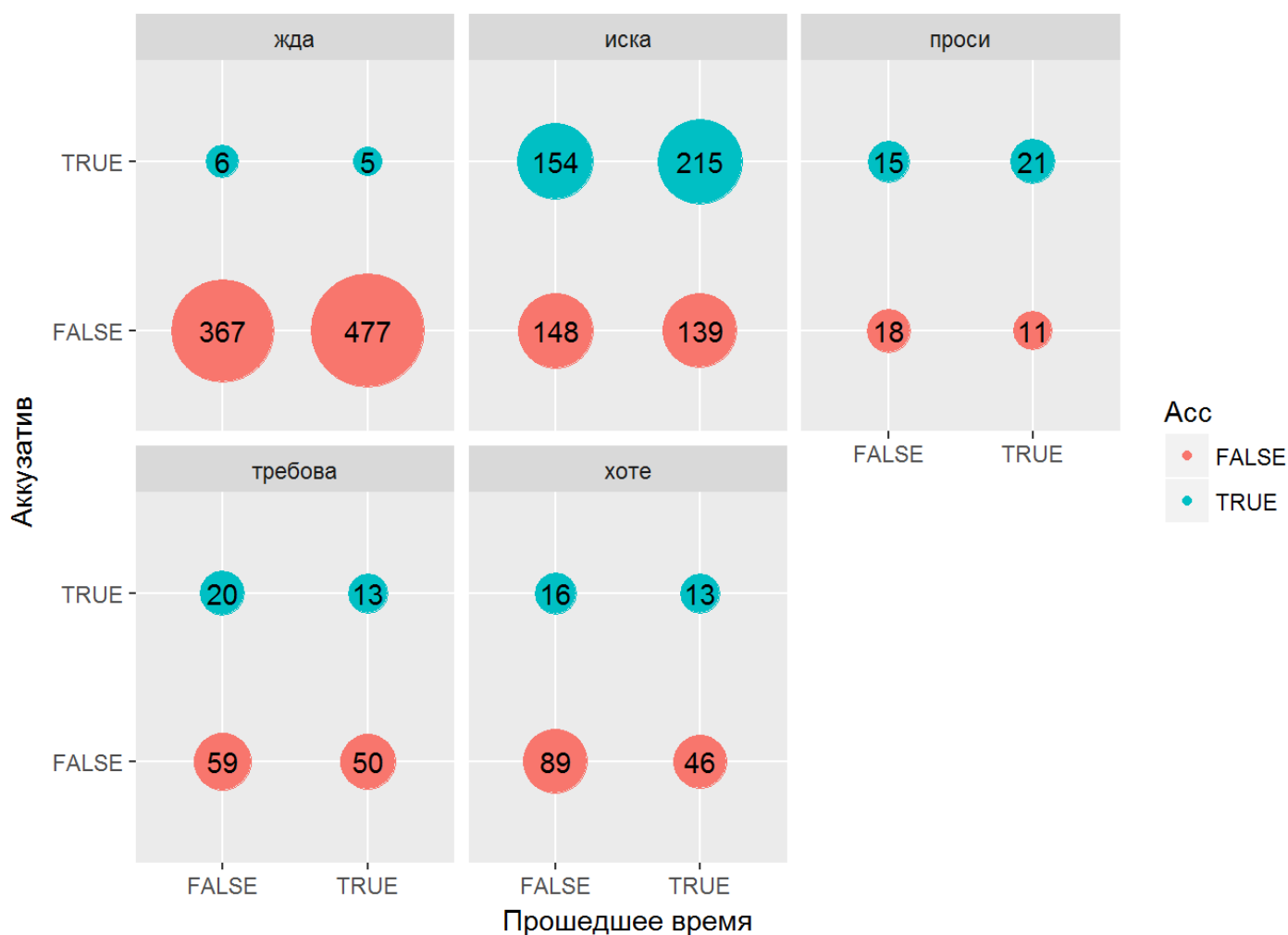
```
data_rel %>%
  ggplot(aes(root, percent, fill = Acc)) +
  geom_bar(stat = "identity") +
  labs(x = "Глагол", y = "") +
  theme_bw() +
  guides(fill = guide_legend(title = "Аккузатив"))
```



Посмотрим на распределение примеров по каждому из факторов по отдельности.

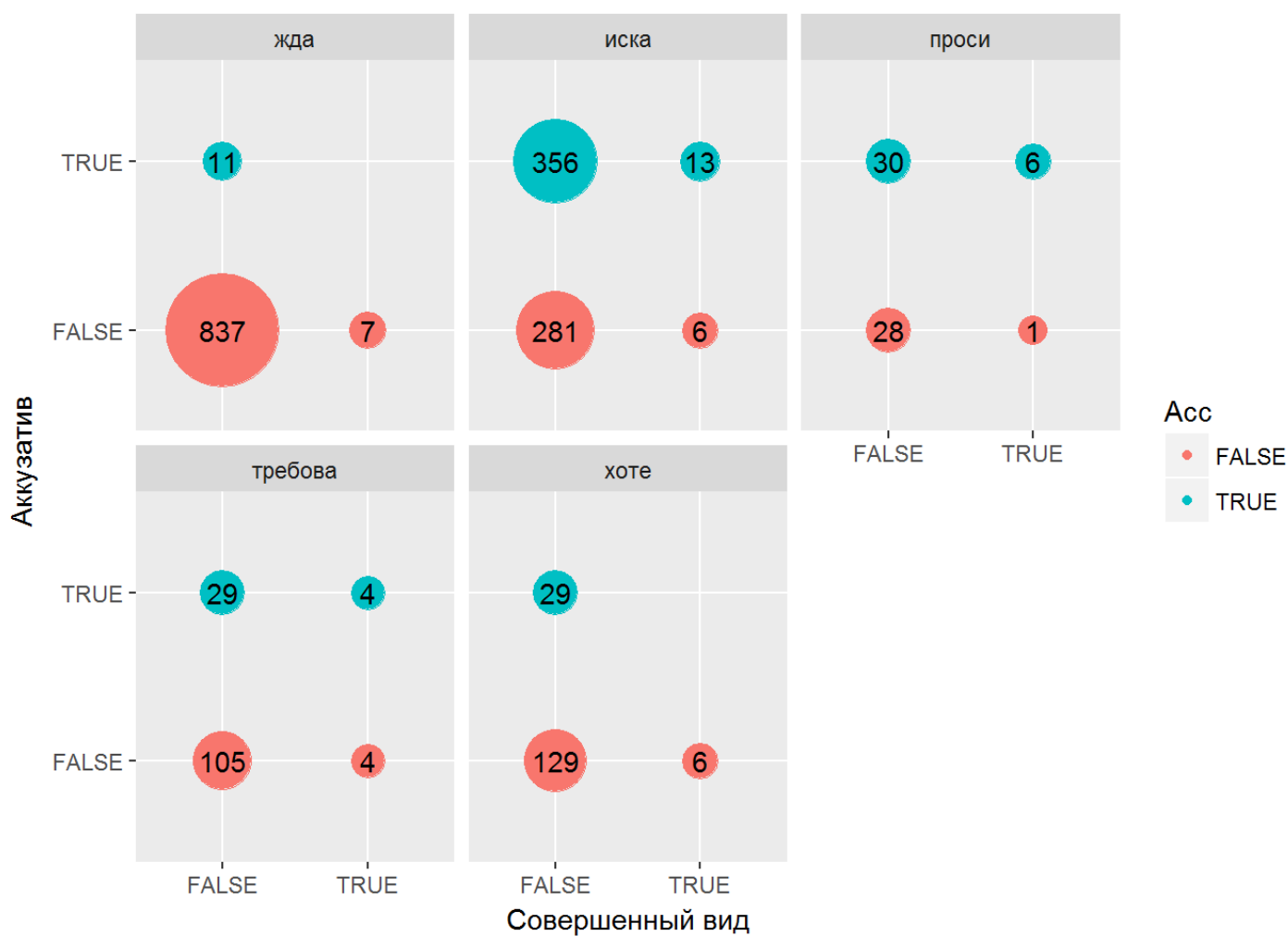
Распределение падежа по фактору времени:

```
data_project %>%
  group_by(past, Acc, root) %>%
  summarise(number = n()) %>%
  ggplot(aes(past, Acc, label = number))+
  geom_point(aes(size = number, color = Acc))+
  geom_text()+
  scale_size(range = c(5, 20))+
  facet_wrap(~root) +
  guides(size = F) +
  labs(x = "Прошедшее время", y = "Аккузатив")
```



Распределение падежа по фактору вида:

```
data_project %>%
  group_by(perfect, Acc, root) %>%
  summarise(number = n()) %>%
  ggplot(aes(perfect, Acc, label = number))+
  geom_point(aes(size = number, color = Acc))+
  geom_text()+
  scale_size(range = c(5, 20))+
  facet_wrap(~root) +
  guides(size = F) +
  labs(x = "Совершенный вид", y = "Аккузатив")
```



Распределение падежа по фактору наличия зависимого прилагательного:

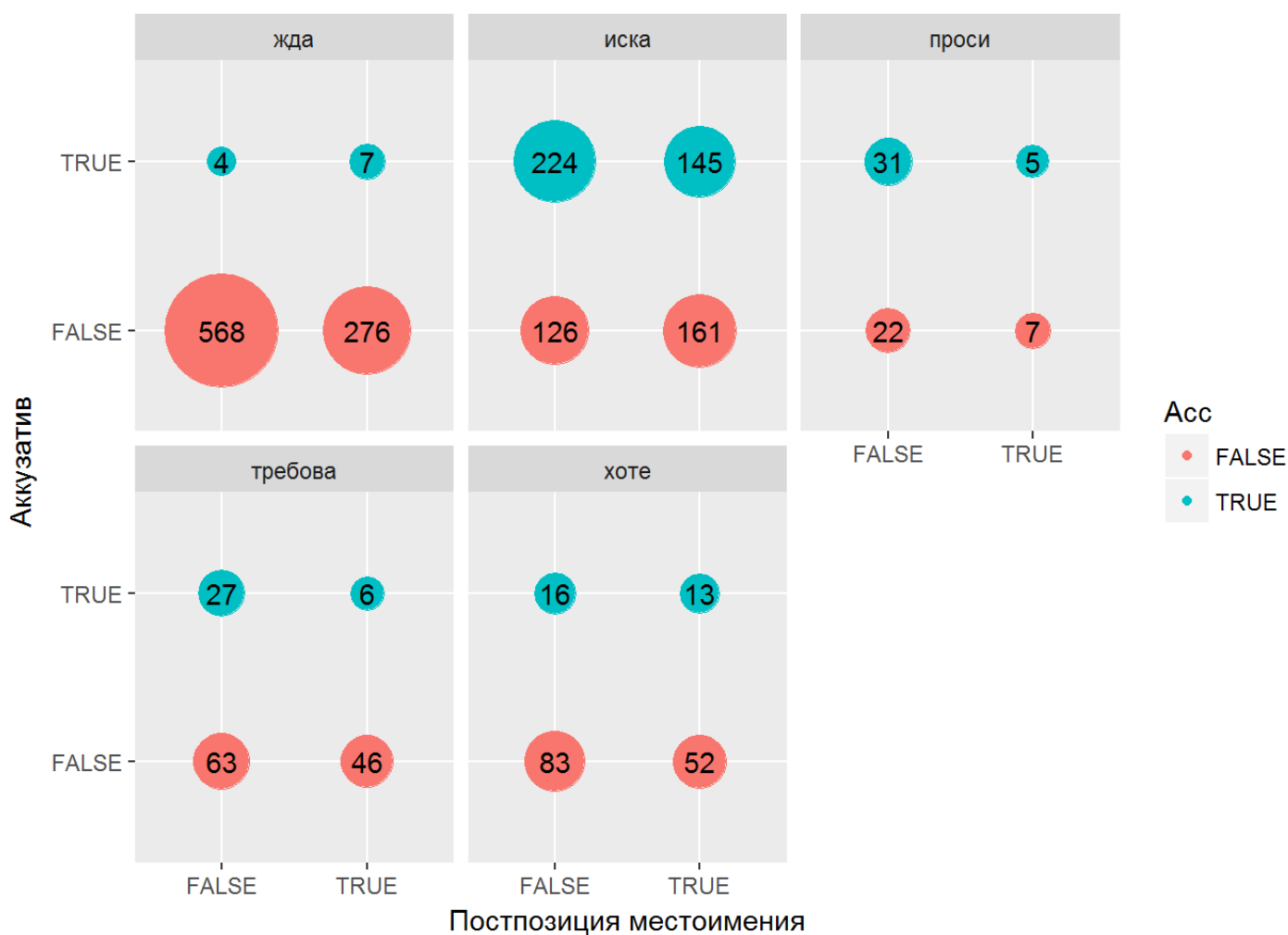
```
data_project %>%
  group_by(adjective, Acc, root) %>%
  summarise(number = n()) %>%
  ggplot(aes(adjective, Acc, label = number))+
  geom_point(aes(size = number, color = Acc))+
  geom_text()+
  scale_size(range = c(5, 20))+
  facet_wrap(~root) +
  guides(size = F) +
  labs(x = "Наличие прилагательного", y = "Аккузатив")
```



Распределение падежа по фактору позиции местоимения по отношению к глаголу:



```
data_project %>%
  group_by(postposition, Acc, root) %>%
  summarise(number = n()) %>%
  ggplot(aes(postposition, Acc, label = number))+
  geom_point(aes(size = number, color = Acc))+
  geom_text()+
  scale_size(range = c(5, 20))+
  facet_wrap(~root) +
  guides(size = F) +
  labs(x = "Постпозиция местоимения", y = "Аккузатив")
```



## Логистическая регрессия

Обучение модели логистической регрессии с использованием всех факторов:

```
data <- read_excel("data_project.xlsx")
data <- data[,1:6]
fit <- glm(Acc ~ ., data = data, family = "binomial")
summary(fit)
```

```
##
## Call:
## glm(formula = Acc ~ ., family = "binomial", data = data)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -1.7612  -0.6392  -0.1584   0.8736   3.1404
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)  -4.371859   0.316931 -13.794 < 2e-16 ***
## past          0.294861   0.132406   2.227 0.025951 *
## perfect       0.680710   0.351473   1.937 0.052778 .
## adjective     0.001277   0.216567   0.006 0.995293
## postposition -0.552085   0.149184  -3.701 0.000215 ***
## rootиска      4.708793   0.316536  14.876 < 2e-16 ***
## rootпроси     4.476417   0.396873  11.279 < 2e-16 ***
## rootтребова   3.180062   0.365358   8.704 < 2e-16 ***
## rootхоте      2.887441   0.369973   7.804 5.98e-15 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##      Null deviance: 2132.9  on 1881  degrees of freedom
## Residual deviance: 1387.6  on 1873  degrees of freedom
## AIC: 1405.6
##
## Number of Fisher Scoring iterations: 7
```

Использование генитивных форм наиболее вероятно без учета других факторов. При этом коэффициенты для разных корней говорят о том, что фактор корня влияет сильнее всего на зависимую переменную. Глагол “ждать” с наибольшей вероятностью выбирает форму родительного падежа, сильно отличаясь от четырех других корней (что соответствует увиденному нами на графиках). Далее с повышением вероятности идут корни *хотеть*, *требовать*, *просить*, *искать*.

## Деревья решений

```
library(party)
```

```
## Warning: package 'party' was built under R version 3.3.3
```

```
## Loading required package: grid
```

```
## Loading required package: mvtnorm
```

```
## Loading required package: modeltools
```

```
## Loading required package: stats4
```

```
## Loading required package: strucchange
```

```
## Warning: package 'strucchange' was built under R version 3.3.3
```

```
## Loading required package: zoo
```

```
## Warning: package 'zoo' was built under R version 3.3.3
```

```
##  
## Attaching package: 'zoo'
```

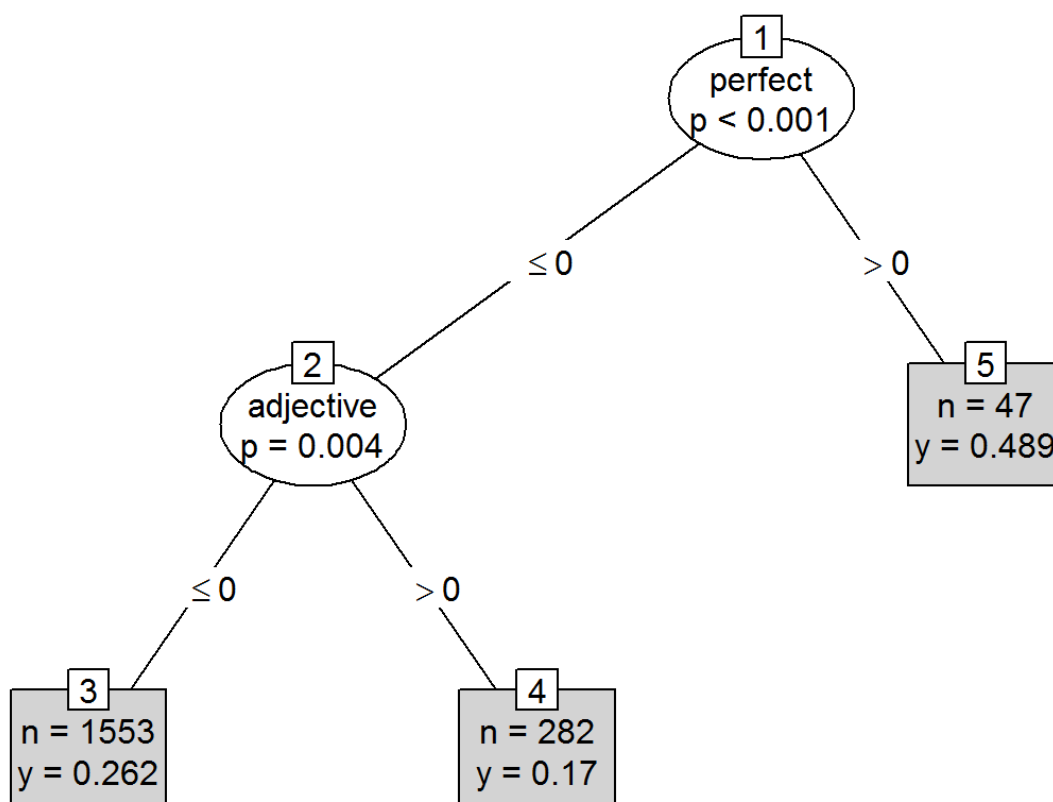
```
## The following objects are masked from 'package:base':  
##  
## as.Date, as.Date.numeric
```

```
## Loading required package: sandwich
```

```
## Warning: package 'sandwich' was built under R version 3.3.3
```

Дерево решений без учета влияния фактора глагольного корня:

```
my_tree <- ctree(Acc ~ past+perfect+adjective+postposition, data = data)  
plot(my_tree, type = "simple")
```



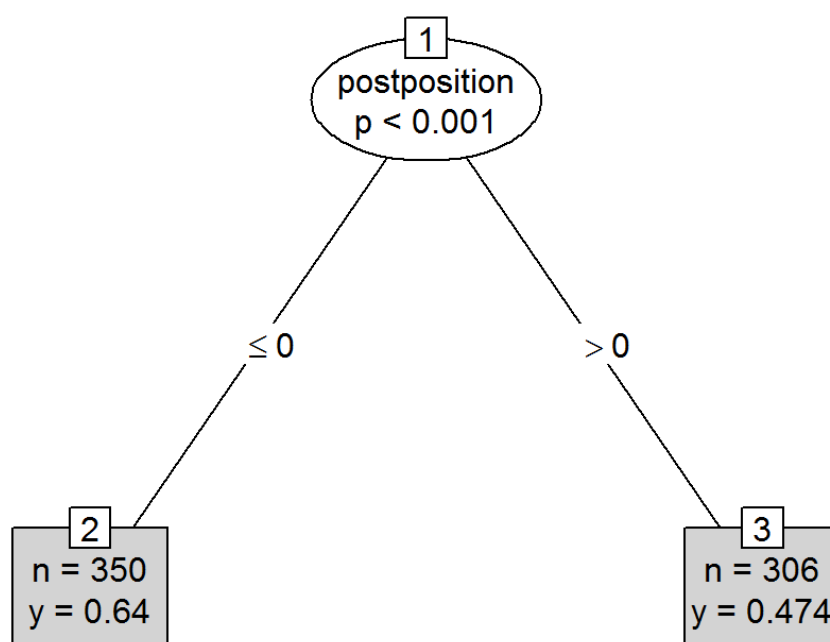
Модель, обученная на данных всех корней, говорит нам о том, что для глаголов совершенного вида падежи почти равновероятны, в то время как для глаголов несовершенного вида более вероятен генитив, особенно при наличии зависимого прилагательного.

Построим деревья решений на наших данных, отдельно для каждого корня.

Дерево решений для глагола “искать”:

```

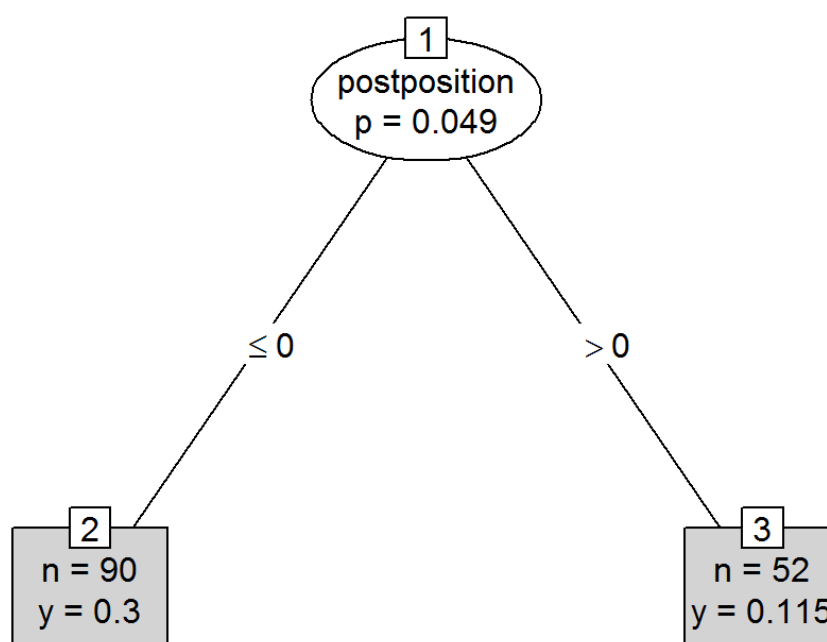
data %>%
  filter(root=='иска') -> data_iska
my_tree <- ctree(Acc ~ past+perfect+adjective+postposition, data = data_iska)
plot(my_tree, type = "simple")
  
```



Дерево решений для глагола “требовать”:

```

data %>%
  filter(root=='требова') -> data_treb
my_tree <- ctree(Acc ~ past+perfect+adjective+postposition, data = data_treb)
plot(my_tree, type = "simple")
  
```



Модели предсказывают зависимость от позиции объекта для корней “искать” и “требовать”: при постпозиции вероятность аккумулятива немного снижается. Для других корней все факторы являются недостаточно влиятельными

## Выводы

Лингвистический анализ: полученные результаты наглядно демонстрируют различие стратегий выбора падежной формы DO для разных глаголов и слабую зависимость выбора от таких факторов как вид, время, прилагательное и позиция. Интерпретируя результаты, скажем, что переходность представляется нам не строго бинарной характеристикой глагола, скорее можно говорить о градуальной шкале от Gen до Acc, и исследуемые нами глаголы выстраиваются на ней в следующем порядке ждать → хотеть → требовать → просить → искать.