

The debitive nominalization in Moksha Mordvin: a case of differential object marking

Aigul Zakirova

This study concerns object marking and object agreement observed with a certain modal form (predicative *-ma* nominalization) in Moksha Mordvin. The direct object of this form can be expressed with a nominative, genitive or non-marked form. Chi-square and Fisher's exact test were used to check the influence of different factors as definiteness of the object, its number and linear position. It turns out that object agreement correlates with case, definiteness and linear position of the object.

Introduction and the problem

Moksha Mordvin is one of the Mordvinic languages (> Finno-Ugric > Uralic), the other being Erzya. Moksha and Erzya together count about 390000 speakers. Moksha is spoken mainly in the Western part of the Republic of Mordovia in European Russia.

Moksha has a rich nominal system: nouns are declined for case, number and definiteness/possessiveness.

In finite clauses Moksha has differential object marking (DOM) and object agreement. Finite verbs can take as complements genitive form (marked for number and definiteness/possessiveness) and the non-marked form, which is only marked for number. Choice of the noun form depends on its definiteness: definite objects are in genitive, indefinite objects are non-marked. Genitive objects always trigger object agreement, unless in some special pragmatic contexts, non-marked forms never do so.

So, the causal link between the object case marking and object agreement is as follows:
DEFINITENESS -> CASE -> AGREEMENT

So most cases of DOM look like that:

DEFINITE -> GENITIVE -> OBJECT AGREEMENT

INDEFINITE -> NON_MARKED -> NO OBJECT AGREEMENT

However, there is a less straightforward case of DOM in a certain modal construction with a nominalization.

Basic facts about the *-ma* form

In general, the suffix *-ma* in Moksha Mordvin is used to derive nominals from verbs. The resulting form has some syntactic properties of a noun, e.g. it may be modified by adjectives and adjoin the definiteness marker *-s'*:

1. viškə mora-ma-s' teǰə-nzə miša-nd-a-j
strong sing-NZR-DEF PRON.DAT-POSS.3SG disturb-IPVF-NPST.3SG
'Loud singing will disturb him/ her.'

The *-ma* form is also used as a predicate to convey the meaning of necessity [Bubrikh 1953]. It is often used to provide instructions:

2. kaftə čast-tə mel'ə šejɛɣ-n'ə s'ej-ənd'-əma-t pr'ɛ-s' pez'-əma
two hour-ABL after hair-DEF.PL comb-IPFV-NZR-PL head-DEF wash-
NZR
'Two hours later one should comb one's hair and wash it.'

The predicative *-ma* form is derived from both transitive and intransitive verbs. The subject of the verb always receives dative marking in the *-ma* form:

3. s'ora-n'ɛ-t'i luvə-ma t'ɛ kn'iga-t'
guy-DIM-DEF.SG.DAT read-NZR this book-DEF.SG.GEN
'The boy has to read this book.'

The object however exhibits DOM. It may be in the nominative (2), genitive (3), or non-marked form (4).

4. mon'-d'ejə-n mora-ma mora
I.OBL-PRON.DAT-POSS.1SG sing-NZR song
'I have to sing a song'.

According to my field data, if the object is a 1st or 2nd person pronoun, it is usually in nominative (which probably means that animateness or topicality influences case marking).

As for the object agreement, personal pronouns and objects marked for definiteness/ possessiveness always trigger agreement on the *-ma* form, as in (5), and with the non-marked object agreement is optional (6):

5. mon'-d'ejə-n *mora-ma/ mora-ma-t mora-t'n'ə
I.OBL-PRON.DAT-POSS.1SG sing-NZR/ sing-NZR-PL song-DEF.PL
'I have to sing the songs'.

6. mon'-d'ejə-n mora-ma/ mora-ma-t mora-t
 I.OBL-PRON.DAT-POSS.1SG sing-NZR/ sing-NZR-PL song-PL
 'I have to sing songs'.

So, the significant difference of this DOM observed in the debitive construction from the DOM in finite clauses is that the object may be coded by both genitive and nominative and the object agreement is at least always possible

It seems that the marking of the direct object and the agreement patterns in the debitive construction depend on quite a few parameters. So far we have mentioned that **definiteness** might influence the agreement of the -ma form. However, it is not clear whether other factors influence the agreement on the -ma- form: the **case** of the direct object, or its **linear position**. Besides, it is not clear whether case assignment depends on any other variables – animacy, number or linear position. The difficulty is that some of these factors may depend on other factors, i.e. variable A may depend on the factor B, and B would depend on C. For example, some pragmatic factor might influence case marking which in turn influences agreement patterns or vice versa.

Hypotheses

Data on DOM in finite sentences and the field data mentioned above point that the causal link also goes in the same direction: CASE -> AGREEMENT.

The hypotheses I would like to test are the following:

- case marking of the object NP depends on its animacy;
- case marking of the object NP depends on its linear position;
- case marking of the object NP depends on its number;
- agreement of the -ma form with the object NP depends on its definiteness;
- agreement of the -ma form with the object NP depends on its case marking (nominative or accusative).
- agreement of the -ma form with the object NP depends on its linear position (before or after the -ma form).

The data I have from the field unfortunately do not show frequencies of some options. So another, secondary goal of my research would be to visualize the frequency of dative omission, percentage of NOM vs GEN object marking observations and similar.

Data

Observations

The data I am going to analyze in this paper comes from the corpus of texts published in the newspaper "Mokshen' pravda" (<http://www.mokshen-pr.ru/> (<http://www.mokshen-pr.ru/>)). I am grateful to Maria Kholodilova and Anastasia Kozhemyakina, who during the fieldwork project extracted these texts and preprocessed them making them easier to work with. I employed search in the Microsoft Word file to extract sentences with the debitive -ma form in the present tense. The queries were as follows: "ман"

(NZR-1SG), “маτ” (NZR-2SG), “ма” (NZR.3SG), “ма́тама” (NZR-1PL), “мата́да” (NZR-2PL), “маτ” (NZR-3PL, coincides with the NZR-2SG form). As the next step I went through the sentences that contained matching strings, translated them into Russian and filtered the examples that actually contained the debitive -ma form. This way I collected the first 199 occurrences of the form in question.

Let us now import the .csv file with the mark-up

```
df <- read.csv2("C://Users/Aigul/Desktop/deb_nom.csv", header=TRUE)
```

Variables

The .csv file we just downloaded contains data about 199 sentences with the *-ma* form. The variables are the following:

TRANS

Transitivity of the verb from which the -ma form is derived.

TR – transitive verb;

INTR – intransitive verb;

TR_NF_SA – transitive matrix verb used with a non-finite complement, e.g. juksn’əms + INF ‘to forget to do something’;

TR_FIN_SA – transitive matrix verb which requires a finite complement, e.g. sodams + što P ‘to know that P’;

TR_ING – ingestive verb (verbs like ‘eat’ and ‘drink’ in Moksha can assign ablative case, ablative objects behave differently from other types of direct objects);

INTR_NF_SA – intransitive matrix verb used with a non-finite complement, e.g. lotkams + INF.ABL ‘to with draw from doing something’;

REFL – reflexive verb;

TR_DUR – a transitive verb with a temporal object, e.g. t’ijəms kaftə šit ‘to do something for two days’.

DO_POS

Type of the direct object.

NP – noun phrase;

PRO – topical noun phrase not expressed;

PRON – pronoun (demonstratives and interrogatives);

QP – quantified phrase;

FIN_SA – finite complement, e.g. a što-clause;

NF_SA – non-finite complement, e.g. an infinitive;

REFL – reflexive pronoun;

NA – not attested (= there is no direct object).

DO_CASE

Case marking on the direct object.

GEN – genitive;

NOM – nominative;

NM – not marked for case;

ABL – ablative;

NA – not attested (=there is no expressed direct object or it cannot be marked for case).

DO_DEF

Formal definiteness/ possessive marking on the direct object.

DEF – definite;

INDEF – indefinite;

POSS – possessive;

NA – not attested (there is no expressed object or it cannot be marked for definiteness).

DO_CONJUNCT

This variable takes positive value if two or more direct objects of a -ma form are conjoined.

MA_CONJUNCT

This variable takes positive value if two or more -ma forms that have a common direct object are conjoined.

DO_PL_REF

This variable shows whether the referent of the object is singular or plural.

SG – singular

PL* – plural

SUBST** – the NP denotes a substance

NA – not attested (there is no expressed object).

DO_PL_NP

This referent shows whether the object NP is grammatically singular or plural.

SG – singular

PL – plural

NA – not attested (there is no expressed object).

MA_PL

This variable takes positive value if the -ma form bears the plural marker. **NO** – no plural marker;

PL – the -ma form bears the plural marker.

LP

This variable describes the linear position of the object in relation to the *-ma* form.

AV – the objects precedes the *-ma* form;

VA – the object comes after the *-ma* form;

NA – not attested (there is no expressed object)

SUBJ

This variable describes the presense of the dative subject

NO – the dative subject is omitted;

DAT – the dative subject is expressed.

ANIM

This variable takes a positive value if the object is animate

Fragment of the mark-up

Let us have a look at the first 10 rows of the dataframe:

```
df_head <- head(df, 10)
df_head
```

##	TRANS	DO_POS	DO_CASE	DO_DEF	DO_CONJUNCT	MA_CONJUNCT	DO_PL_REF	DO_PL_NP
## 1	TR	NP	NM	INDEF	NO	NO	SG	SG
## 2	INTR	<NA>	<NA>	<NA>	<NA>	NO	<NA>	<NA>
## 3	TR	NP	NOM	DEF	NO	NO	PL	PL
## 4	INTR	<NA>	<NA>	<NA>	<NA>	NO	<NA>	<NA>
## 5	TR	PRO	<NA>	<NA>	NO	NO	PL	<NA>
## 6	TR	NP	NM	INDEF	NO	NO	PL	PL
## 7	TR	NP	NM	INDEF	NO	NO	SG	SG
## 8	TR	PRON	NOM	<NA>	NO	NO	SG	SG
## 9	TR	NP	NM	INDEF	NO	NO	SG	SG
## 10	INTR	<NA>	<NA>	<NA>	<NA>	NO	<NA>	<NA>

##	MA_PL	LP	SUBJ	ANIM
## 1	NO	VA	NO	NO
## 2	NO	<NA>	NO	<NA>
## 3	PL	VA	NO	NO
## 4	NO	<NA>	NO	<NA>
## 5	PL	<NA>	NO	NO
## 6	NO	VA	NO	NO
## 7	NO	VA	NO	NO
## 8	NO	VA	NO	NO
## 9	NO	VA	NO	NO
## 10	NO	<NA>	NO	<NA>

For example, the 3rd observation is the following sentence:

6. Марстонь вийса ванфтомат велетне, вдь тоса «эряхть» юроньке, коеньке-обуцяньке.

marst-ən' vij-sə van-ft-əma-t **vel'ə-t'n'ə** vəd' tosə er'ε-j̥-t' jurən'kə
kojən'kə obuc'εn'kə

common-GEN strength-INSTR look-CAUS-NZR-PL **village-DEF.PL**
PTCL there live-NPST.3-PL root-1PL.POSS tradition-1PL.POSS
custom-1PL.POSS

'Villages must be cared for by all of us because there our roots are,
there our customs live'.

TRANS: the verb *vanftəms* is transitive, so the variable takes the value **TR**;

DO_POS: the verb has the direct object *vel'ət'n'ə* which is an NP, so the variable takes the value **NP**;

DO_CASE: *vel'ət'n'ə* is in nominative, so the variable takes the value **NOM**;

DO_DEF: *vel'ət'n'ə* is definite so the variable takes the value **DEF**;

DO_CONJUNCT: *vel'ət'n'ə* is not conjoined with any other constituent, so the variable takes the value **NO**;

MA_CONJUNCT: *vanftəmat* is not conjoined with any other constituent, so the variable takes the value **NP**;

DO_PL_REF: the referent of the NP *vel'ət'n'ə* is plural so the variable takes the value **PL**;

DO_PL_NP: the NP *vel'ət'n'ə* is grammatically plural so the variable takes the value **PL**;

MA_PL: the *-ma* form is marked for plural, so the variable takes the value **PL**;

LP: the linear position of the object is after the *-ma* form, so the variable takes the value **VA**;

SUBJ: the dative subject is not expressed, so the variable takes the value **NO**;

ANIM: the object is inanimate, so the variable takes the value **NO**

Let us now have a look at the summary of the data:

```
summary(df)
```

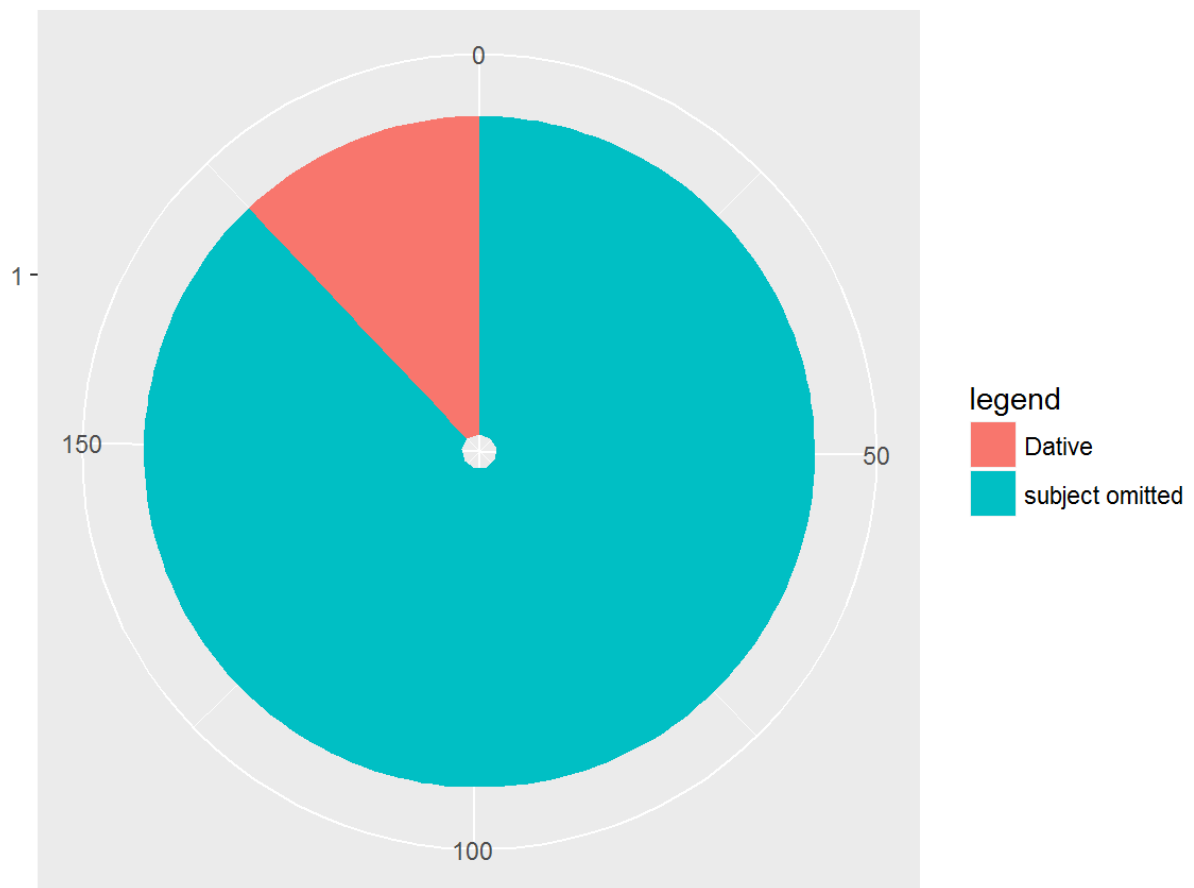
```
##          TRANS          DO_POS    DO_CASE    DO_DEF    DO_CONJUNCT
## TR          :145    NP          :119    ABL : 5    DEF :60    NO :157
## INTR        : 31    PRON        : 12    GEN :21    INDEF:64    YES : 9
## TR_ING      : 10    QP          : 12    NM :50    POSS :12    NA's: 33
## TR_FIN_SA   : 7    PRO          : 10    NOM :59    NA's :63
## INTR_NF_SA: 2    FIN_SA : 8    NA's:64
## REFL        : 2    (Other): 4
## (Other)     : 2    NA's : 34
## MA_CONJUNCT DO_PL_REF DO_PL_NP  MA_PL    LP        SUBJ        ANIM
## NO :191     PL :68    PL :58    NO:152    AV :66    DAT: 24    NO :158
## YES: 8      SG :75    SG :84    PL: 47    VA :91    NO :175    YES : 6
##           SUBST:11    NA's:57    NA's:42    NA's: 35
##           NA's :45
##
##
##
##
```

Some visualization

The dative argument is often omitted. The following piechart demonstrates it:

```
pie <- ggplot(df, aes(x = factor(1), fill = factor(df$SUBJ))) + geom_bar()
pie + coord_polar(theta = "y") + ggtitle("Subject expression") + xlab("") + ylab("") + scale_fill_discrete(name = "legend", labels = c("Dative", "subject omitted"))
```


Subject expression

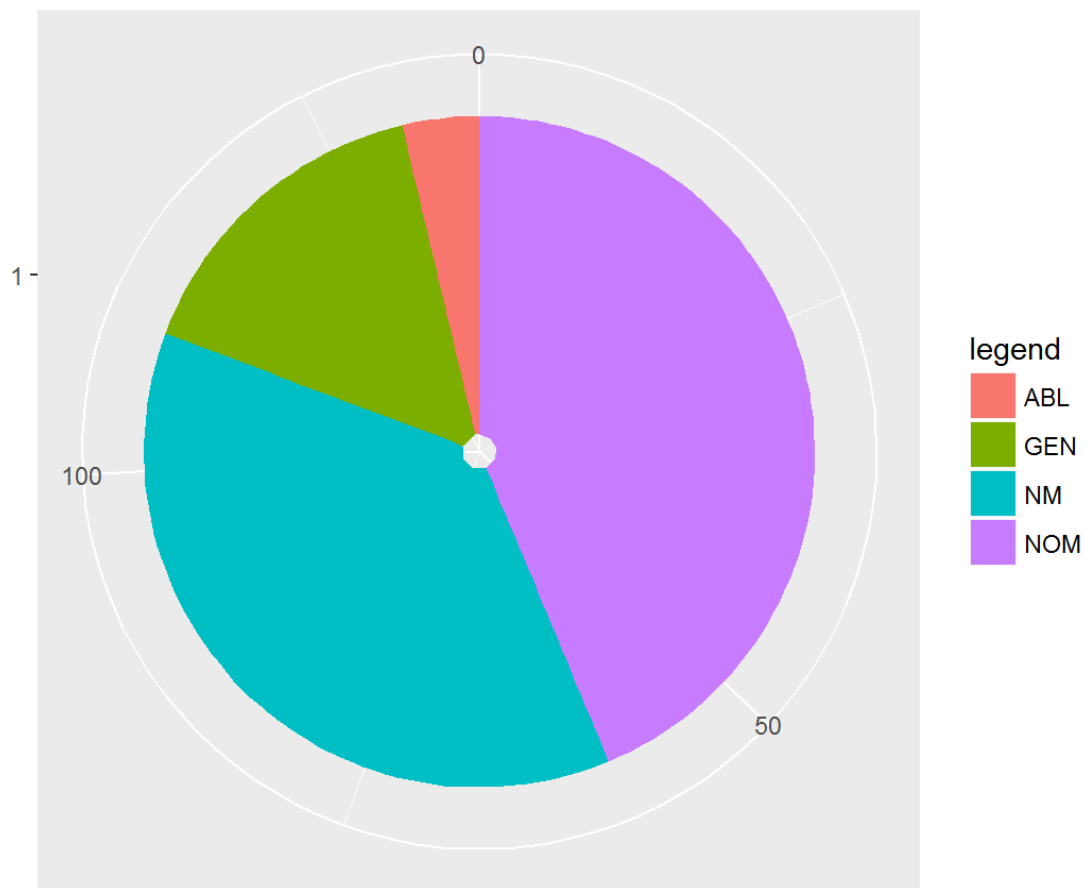


Now we will visualize the distribution of object marking strategies. Let us subset the dataframe so that we do not include NA into the visualization:

```
df_subs_na <- subset(df, df$DO_CASE != "NA")

pie <- ggplot(df_subs_na, aes(x = factor(1), fill = factor(df_subs_na$DO_CASE))) +
  geom_bar()
pie + coord_polar(theta = "y") + ggtitle("Object marking") + xlab("") + ylab("") +
  scale_fill_discrete(name = "legend")
```

Object marking



```
df_subs <- subset(df, df$DO_CASE == "NOM" | df$DO_CASE == "GEN")
```

Testing hypotheses

Case assignment

Does case marking of the object NP depend on its animacy?

As we can see, there are only 6 animate objects, so let us use Fisher's exact test.

The **null hypothesis: case marking and animacy of the object are independent.**

```
anim_table <- table(df$DO_CASE, df$ANIM)
anim_table
```

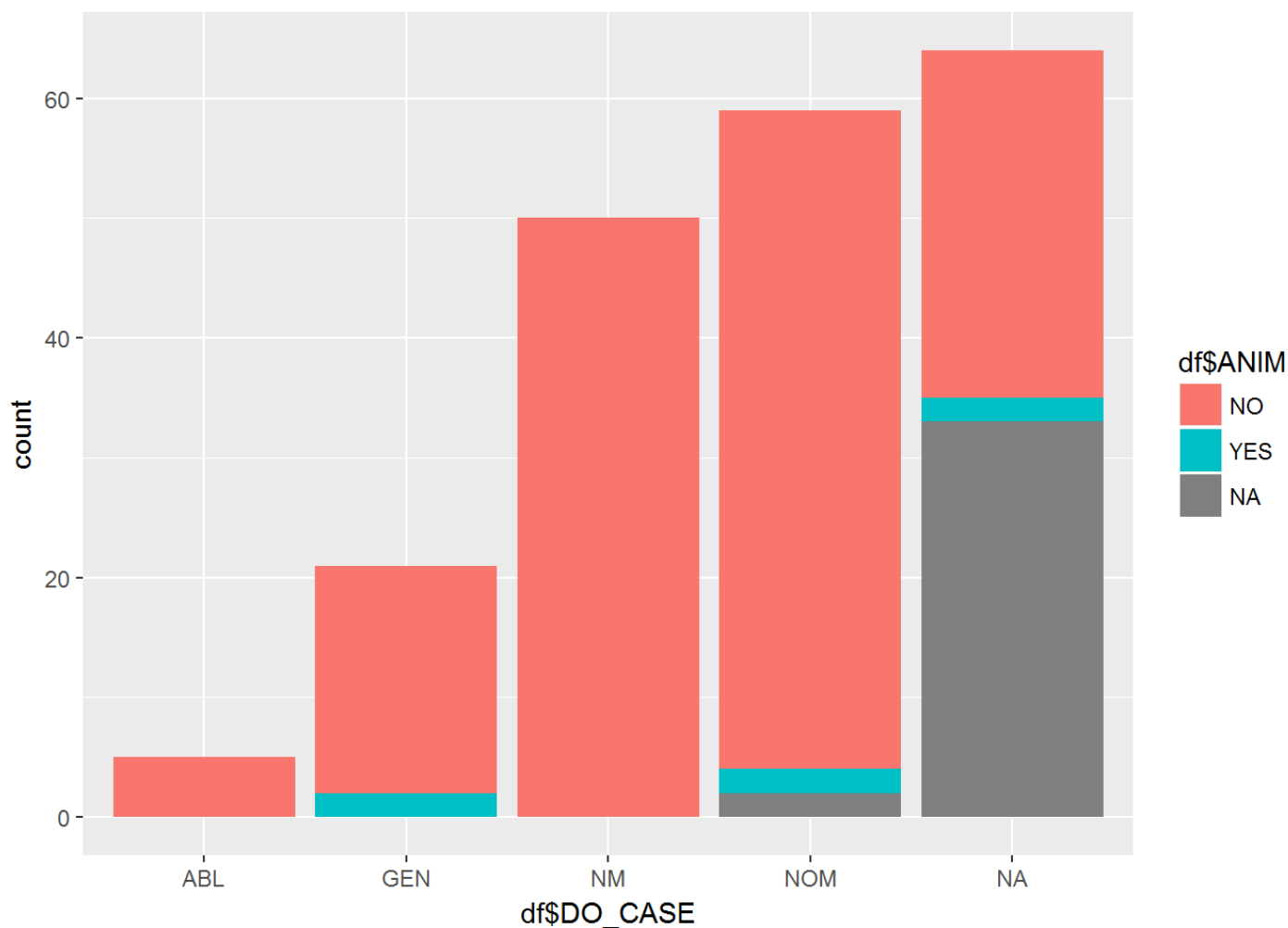
```
##
##          NO YES
##  ABL    5   0
##  GEN   19   2
##  NM    50   0
##  NOM   55   2
```

```
fisher.test(anim_table)
```

```
##
## Fisher's Exact Test for Count Data
##
## data:  anim_table
## p-value = 0.1615
## alternative hypothesis: two.sided
```

According to the Fisher's test, case marking and animacy of the object are independent. As the variables are categorical, it seems reasonable to make a bar chart. The following one shows the occurrences of animate objects.

```
ggplot(df, aes(df$DO_CASE)) + geom_bar(aes(fill = df$ANIM))
```



In this test we dealt with variables that had multiple values. However, what if we subset the dataframe we had so that its subset only contains DOs in nominative and genitive?

```
df_subs <- subset(df, df$DO_CASE == "NOM" | df$DO_CASE == "GEN")
```

```
anim_table_subs <- table(df_subs$DO_CASE, df_subs$ANIM)
anim_table_subs
```

```
##
##      NO YES
##  ABL  0   0
##  GEN 19   2
##  NM   0   0
##  NOM 55   2
```

```
fisher.test(anim_table_subs)
```

```
##
## Fisher's Exact Test for Count Data
##
## data:  anim_table_subs
## p-value = 0.2923
## alternative hypothesis: two.sided
```

This data also does not show any correlation between case marking and animacy.

Does case marking of the object NP depend on its linear position?

Let us create a contingency table.

```
cm_lin_table <- table(df$DO_CASE, df$LP)
cm_lin_table
```

```
##
##      AV VA
## ABL  1  4
## GEN 10 11
## NM   19 31
## NOM  36 23
```

Now we'll apply the chi-square test. **Our hypothesis is that case marking and linear position are independent.**

```
chisq.test(cm_lin_table)
```

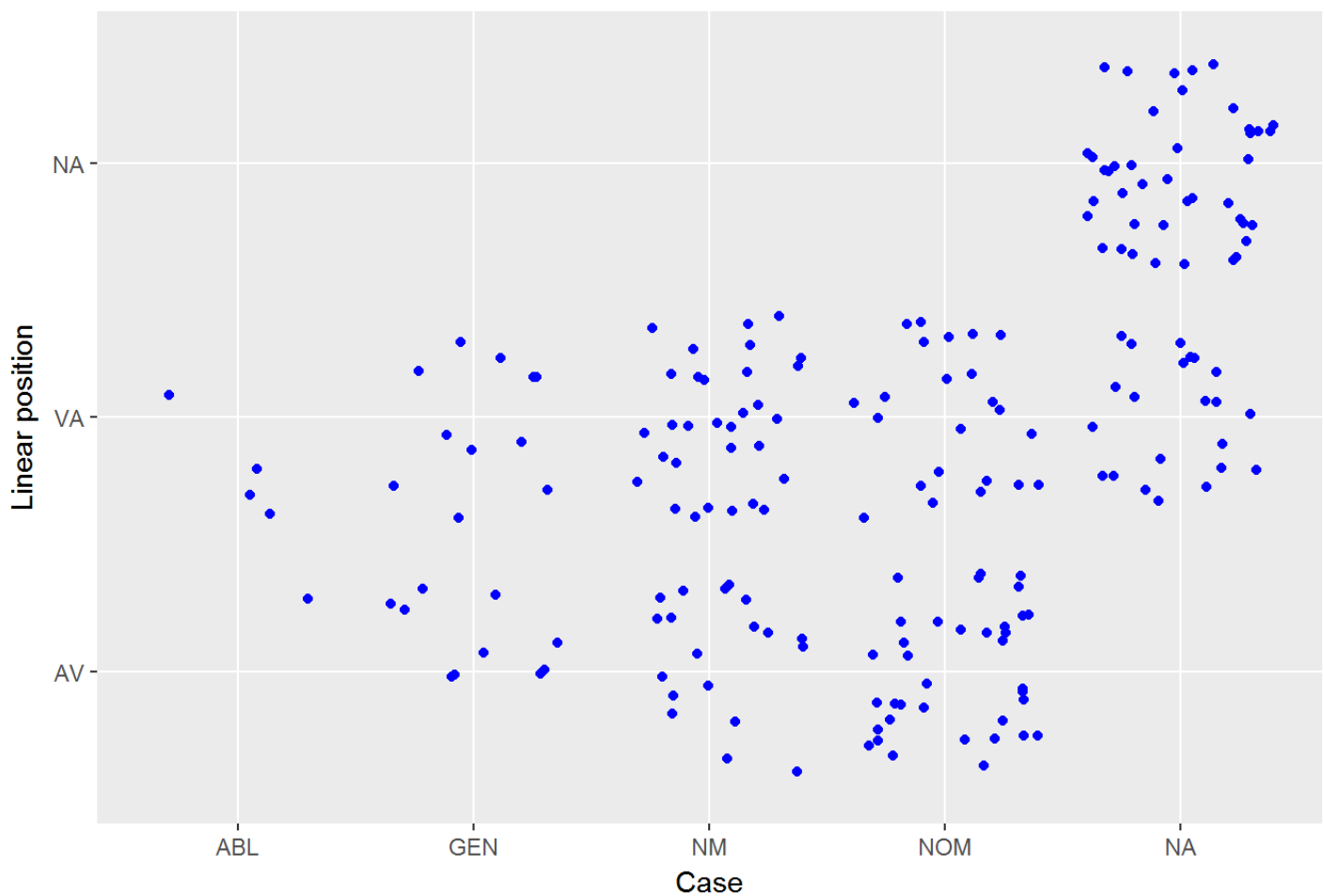
```
## Warning in chisq.test(cm_lin_table): Chi-squared approximation may be
## incorrect
```

```
##
## Pearson's Chi-squared test
##
## data:  cm_lin_table
## X-squared = 7.5291, df = 3, p-value = 0.05682
```

As the p-value > 0.05, we conclude that these variables are again independent. Now let's create a plot that shows theoretically possible combinations of case and linear position:

```
ggplot(df, aes(x = df$DO_CASE, y = df$LP)) + geom_jitter(color = "blue") + ggtitle
("Case and linear position") + xlab("Case") + ylab("Linear position")
```

Case and linear position



Let us also apply the test to the subset we created:

```
cm_lin_table_subs <- table(df_subs$DO_CASE, df_subs$LP)
cm_lin_table_subs
```

```
##
##      AV VA
##  ABL  0  0
##  GEN 10 11
##  NM   0  0
##  NOM 36 23
```

```
chisq.test(cm_lin_table)
```

```
## Warning in chisq.test(cm_lin_table): Chi-squared approximation may be
## incorrect
```

```
##
## Pearson's Chi-squared test
##
## data:  cm_lin_table
## X-squared = 7.5291, df = 3, p-value = 0.05682
```

Case and linear order also seem independent in this case.

Does case marking of the object NP depend on number?

We will again use the chi-square. The null hypothesis: **case marking and number are independent**.

```
case_num_table <- table(df$DO_CASE, df$DO_PL_NP)
case_num_table
```

```
##
##      PL SG
## ABL  0  5
## GEN 11 10
## NM   14 34
## NOM 26 31
```

```
chisq.test(case_num_table)
```

```
## Warning in chisq.test(case_num_table): Chi-squared approximation may be
## incorrect
```

```
##
## Pearson's Chi-squared test
##
## data:  case_num_table
## X-squared = 7.781, df = 3, p-value = 0.05076
```

```
case_num_table_subs <- table(df_subs$DO_CASE, df_subs$DO_PL_NP)
case_num_table_subs
```

```
##
##      PL SG
## ABL  0  0
## GEN 11 10
## NM   0  0
## NOM 26 31
```

```
chisq.test(case_num_table_subs)
```

```
## Warning in chisq.test(case_num_table_subs): Chi-squared approximation may
## be incorrect
```

```
##
## Pearson's Chi-squared test
##
## data: case_num_table_subs
## X-squared = NaN, df = 3, p-value = NA
```

In both cases the variables seem independent.

Conclusion

We have examined different factors that could potentially influence the case assignment in the debitive construction: animacy, linear position and number. We have not found any statistically significant dependencies between case assignment and any of these factors, so the choice between nominative and accusative remains to be explained.

Agreement

Does agreement of the -ma form with the object NP depend on its definiteness?

First, let us subset the dataframe so that we only have NPs denoting plural entities:

```
agr_subset <- subset(df, df$DO_PL_REF == "PL")
```

Let us create a contingency table.

```
def_agr_table <- table(agr_subset$DO_DEF, agr_subset$MA_PL)
def_agr_table
```

```
##
##      NO PL
## DEF    2 31
## INDEF 17  2
## POSS   1  5
```

Let us use the chi-square test. Our **null hypothesis is that the definiteness of the NP and agreement morphology on the -ma form are independent.**

```
chisq.test(def_agr_table)
```



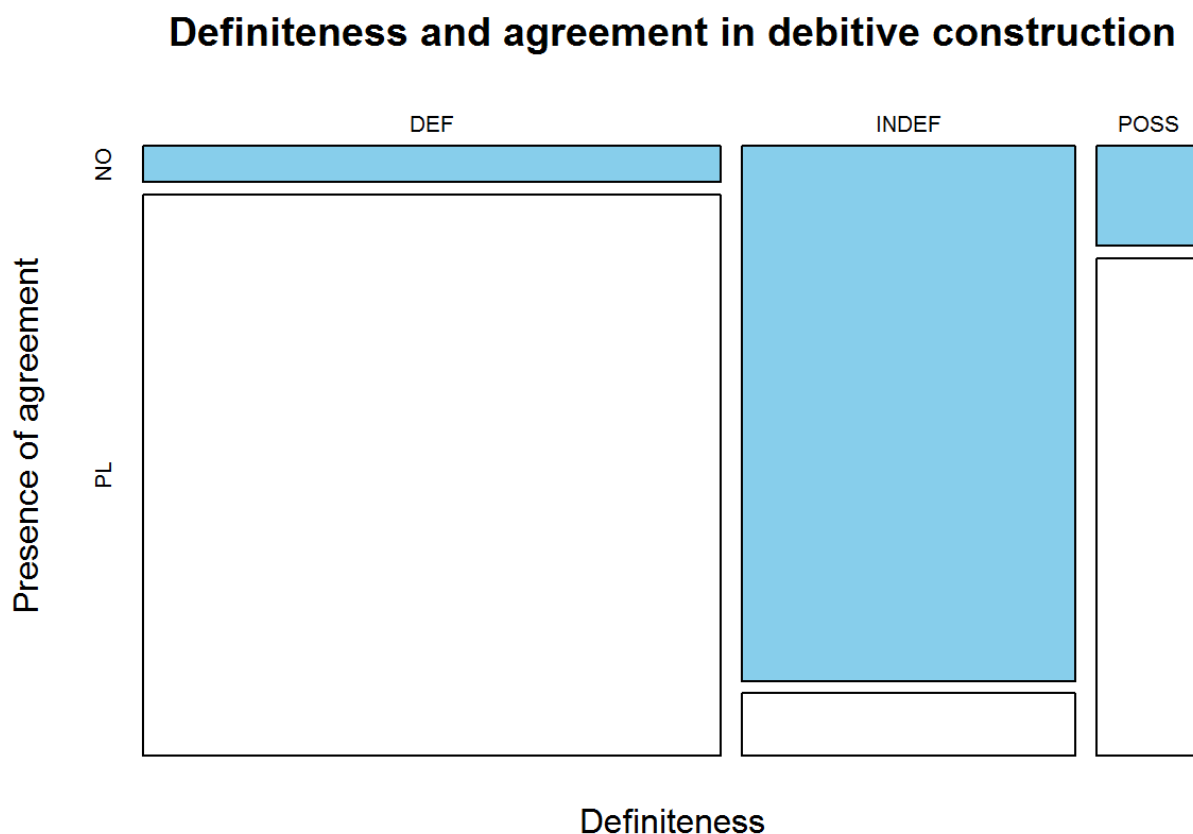
```
## Warning in chisq.test(defAgrTable): Chi-squared approximation may be
## incorrect
```

```
##
## Pearson's Chi-squared test
##
## data:  defAgrTable
## X-squared = 38.075, df = 2, p-value = 5.398e-09
```

This chi-square test's result shows us that the null hypothesis should be dismissed: $p\text{-value} < 0.05$. So, one of our hypotheses was correct: agreement with the object depends on its definiteness.

Now we will visualize the relationship between these variables with a mosaic plot.

```
dfTab <- table(agr_subset$DO_DEF, agr_subset$MA_PL)
mosaicplot(dfTab, main = "Definiteness and agreement in debitive construction", x
lab = "Definiteness", ylab = "Presence of agreement", color = c("skyblue", "white"
))
```



Does agreement of the -ma form with the object NP depend on its case marking (nominative or accusative)?

Let us create a contingency table.

```
caseAgrTable <- table(agr_subset$DO_CASE, agr_subset$MA_PL)
caseAgrTable
```

```
##
##           NO PL
##  ABL    0  0
##  GEN    2  9
##  NM   15  1
##  NOM    0 28
```

We will conduct Fisher's test. The **null hypothesis is that case marking and agreement are independent.**

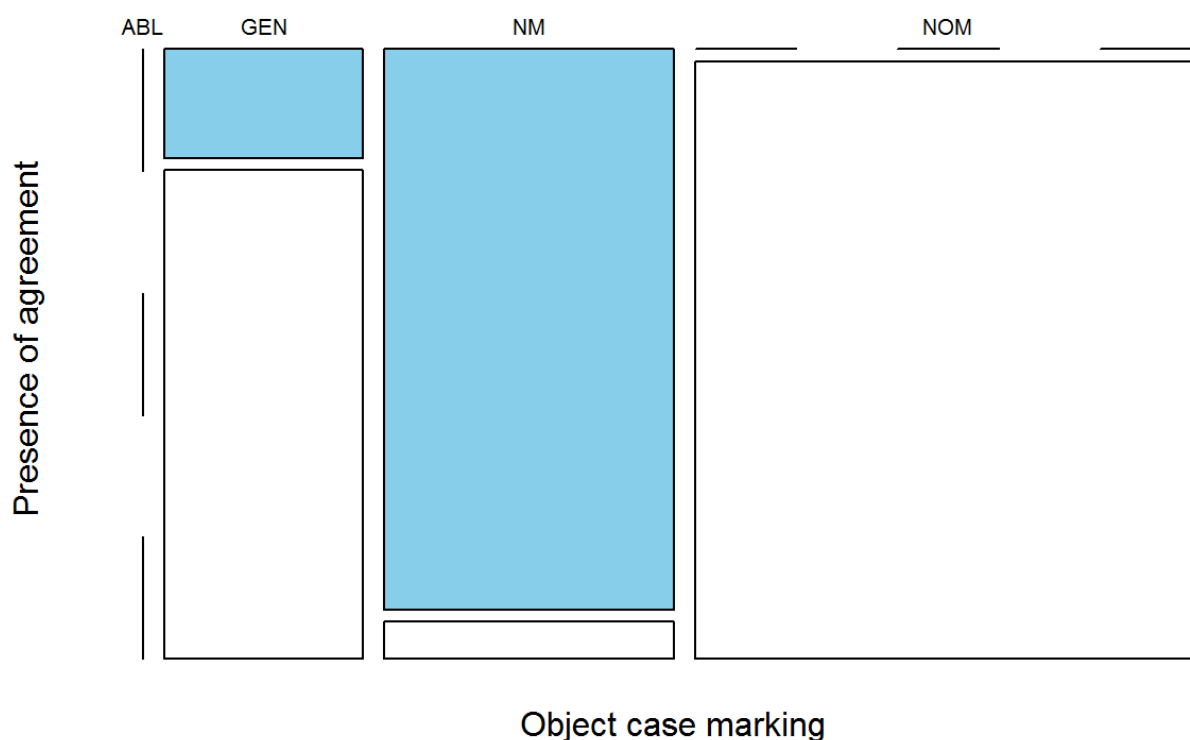
```
fisher.test(caseAgrTable)
```

```
##
## Fisher's Exact Test for Count Data
##
## data:  caseAgrTable
## p-value = 1.347e-11
## alternative hypothesis: two.sided
```

Again, the null hypothesis is dismissed: case marking and agreement seem to be dependent.

```
dfTab2 <- table(agr_subset$DO_CASE, agr_subset$MA_PL)
mosaicplot(dfTab2, main = "Object case marking and agreement in debitive construc
tion", xlab = "Object case marking", ylab = "Presence of agreement", color = c("sk
yblue", "white"))
```

Object case marking and agreement in debitive construction



Does agreement of the -ma form with the object NP depend on their linear position?

```
lp_agr_table <- table(agr_subset$LP, agr_subset$MA_PL)
lp_agr_table
```

```
##
##      NO PL
##  AV   3 22
##  VA  18 18
```

Now we will apply the chi-square test. The **null hypothesis is that linear position and agreement are independent.**

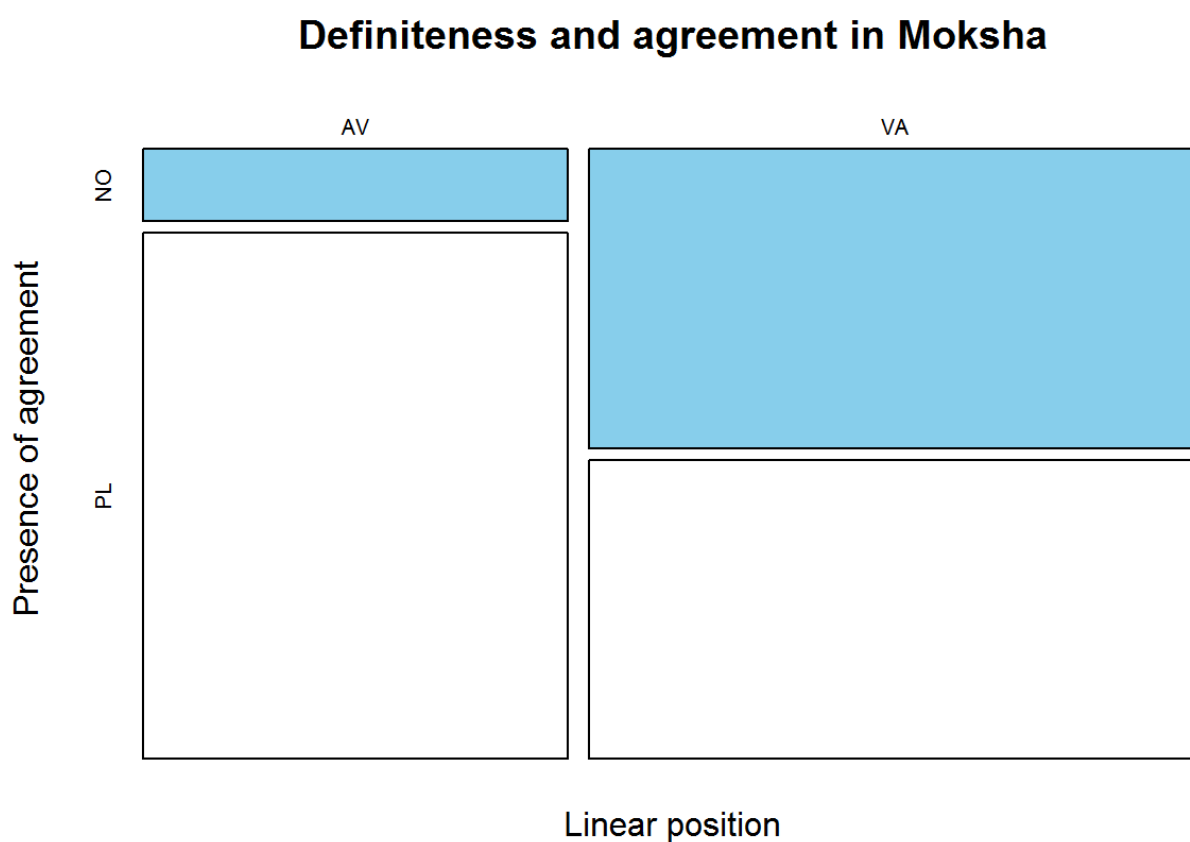
```
chisq.test(def_agr_table)
```

```
## Warning in chisq.test(def_agr_table): Chi-squared approximation may be
## incorrect
```

```
##
## Pearson's Chi-squared test
##
## data:  def_agr_table
## X-squared = 38.075, df = 2, p-value = 5.398e-09
```

Here again, the null hypothesis is dismissed: agreement is influenced by the linear position.

```
df_tab3 <- table(agr_subset$LP, agr_subset$MA_PL)
mosaicplot(df_tab3, main = "Definiteness and agreement in Moksha", xlab = "Linear
position", ylab = "Presence of agreement", color = c("skyblue", "white"))
```



Conclusion

All the factors that we examined – definiteness, case and linear order – influence object agreement with the *-ma* form. **Definite object**, **nominative marking** and **preposition of the object** make the probability of agreement higher.

Results obtained

The results of this study are the following: case and agreement are dependent on each other. Besides, definiteness and linear position influence the agreement.

So, we could propose the following model:

CASE

LINEAR POSITION -> AGREEMENT

DEFINITENESS

Or the following:

LINEAR POSITION -> AGREEMENT -> CASE

DEFINITENESS

The first one seems more plausible if one takes into account the DOM the occurs in finite clauses. Further investigation is needed to shed light on the factors that govern the NOM and GEN assignment.

Sources

Bubrikh, D.V. Istoricheskaya grammatika erzyanskogo yazyka. The historical grammar of Erzya. Saransk, 1953

RStudio. Version 1.0.136 – © 2009-2016 RStudio, Inc.

tidyverse-package: <https://cran.r-project.org/web/packages/tidyverse/index.html> (<https://cran.r-project.org/web/packages/tidyverse/index.html>)

vcd package: <https://cran.r-project.org/web/packages/vcd/index.html> (<https://cran.r-project.org/web/packages/vcd/index.html>)

Mokshen' pravda: <http://www.mokshen-pr.ru/> (<http://www.mokshen-pr.ru/>)