

Supplementary information II: Exploratory analysis

Januar 15, 2024

Contents

1	Load libraries	2
2	Import data	2
3	Prepare data	2
3.1	Add column Vpos	2
3.2	Add column AO	2
3.3	Main word order patterns	4
4	H1: Distance between connective and finite verb	5
5	H2: Position of elements	8
6	H3: Position of finite verb	10
7	H4: Development of patterns	11
8	H5: Frequency biases with CPTs	12
9	H6: Frequency shifts for CTPs	13
10	References	13

1 Load libraries

```
library(tidyr)
library(stringr)
library(dplyr)
library(ggplot2)
library("knitr")
library(cluster)
library(ggpubr)
```

2 Import data

We first import the data frame containing our 875 observations (for sampling see Supplementary Information I).

```
# Read dataframe, remove first column
all.df <- readr::read_csv("all-df.csv")
all.df <- all.df[, -1]
```

3 Prepare data

Next, we merge the two values C and I for Vfin_position and add a new variable Vpos. This variable will henceforth serve as a basis for exploring the relevance of the position of the finite verb.

3.1 Add column Vpos

```
all.df %>% mutate(., Vpos = case_when(Vfin_position == "C" |
                                     Vfin_position == "I" ~ "CI",
                                     Vfin_position == "M" ~ "M",
                                     Vfin_position == "F" ~ "F",
                                     Vfin_position == "dna" ~ "dna")
) -> all.ci
```

3.2 Add column AO

In order to be able to identify possible patterns for the ordering of arguments in the clauses introduced by COMP, DIR, COMP-DIR, we simplify the variable Vfin_argument and add a new variable AO.

```
# Simplify argument order

## remove p from arguments
AO.df <- all.df
AO.df$AO <- all.df$Vfin_arg
AO.df$AO <- gsub("[a-z]", "", AO.df$AO)

## filter combinations of Argument Order and clean data

## V, DO, S: n=121
DOS <- AO.df %>% filter(., grepl("DO", AO.df$AO) & grepl("S", AO.df$AO) &
                        !grepl("IO", AO.df$AO) & !grepl("OO", AO.df$AO))
DOS$AO[DOS$AO == "S_S_V__DO"] <- "S_S_V_DO"

## V, IO, S: n=18
```

```

IOS <- AO.df %>% filter (., !grepl("DO", AO.df$AO) & grepl("S", AO.df$AO) &
                        grepl("IO", AO.df$AO) & !grepl("OO", AO.df$AO))

## V, OO, S: n=74
OOS <- AO.df %>% filter (., !grepl("DO", AO.df$AO) & grepl("S", AO.df$AO) &
                        !grepl("IO", AO.df$AO) & grepl("OO", AO.df$AO))
OOS$AO[OOS$AO == "OO_V_S"] <- "OO_V_S"
OOS$AO[OOS$AO == "S_V_OO"] <- "S_V_OO"

## V, DO, IO, S: n=22
DOIOS <- AO.df %>% filter (., grepl("DO", AO.df$AO) & grepl("S", AO.df$AO) &
                        grepl("IO", AO.df$AO) & !grepl("OO", AO.df$AO) )

## V, DO, IO, OO, S: n=1
DOIOOOS <- AO.df %>% filter (., grepl("DO", AO.df$AO) & grepl("S", AO.df$AO) &
                        grepl("IO", AO.df$AO) & grepl("OO", AO.df$AO) )

## V, DO, IO: n=26
DOIO <- AO.df %>% filter (., grepl("DO", AO.df$AO) & !grepl("S", AO.df$AO) &
                        grepl("IO", AO.df$AO) & !grepl("OO", AO.df$AO) )
DOIO$AO[DOIO$AO=="IO_DO_V"] <- "IO_DO_V"
DOIO$AO[DOIO$AO=="DO_(IO)_V"] <- "DO_IO_V"

## V: n=99 (Vdna includes 46 "dna", they are excluded for "V")
Vdna <- AO.df %>% filter (., !grepl("DO", AO.df$AO) & !grepl("S", AO.df$AO) &
                        !grepl("IO", AO.df$AO) & !grepl("OO", AO.df$AO) )
V <- Vdna %>% filter (., AO != "")

## V, DO, OO: N=36
DOOO <- AO.df %>% filter (., grepl("DO", AO.df$AO) & !grepl("S", AO.df$AO) &
                        !grepl("IO", AO.df$AO) & grepl("OO", AO.df$AO) )

## V, DO: n=139
DO <- AO.df %>% filter (., grepl("DO", AO.df$AO) & !grepl("S", AO.df$AO) &
                        !grepl("IO", AO.df$AO) & !grepl("OO", AO.df$AO) )
DO$AO[DO$AO=="DO_V"] <- "DO_V"

## V, IO: n=6
IO <- AO.df %>% filter (., !grepl("DO", AO.df$AO) & !grepl("S", AO.df$AO) &
                        grepl("IO", AO.df$AO) & !grepl("OO", AO.df$AO) )

## V, S: 220
S <- AO.df %>% filter (., !grepl("DO", AO.df$AO) & grepl("S", AO.df$AO) &
                        !grepl("IO", AO.df$AO) & !grepl("OO", AO.df$AO))
S$AO[S$AO=="S_V"] <- "S_V"
S$AO[S$AO=="S_V_S2"] <- "S_V_S2"
S$AO[S$AO=="S_V_D2"] <- "S_V_S2"
S$AO[S$AO=="VS2"] <- "V_S2"
S$AO[S$AO=="V_S_S2"] <- "V_S_S2"
S$AO[S$AO=="S_P_S2"] <- "S_V_S2"

## V, OO: n=55
OO <- AO.df %>% filter (., !grepl("DO", AO.df$AO) & !grepl("S", AO.df$AO) &

```

```
!grepl("IO", AO.df$AO) & grepl("OO", AO.df$AO))
```

3.3 Main word order patterns

We identify the main word order patterns and divide them into those with and those without overt S(subject).

```
## Patterns with S: WOS.df
```

```
WOS.df <- bind_rows(DOS, S)
```

```
WOS.df %>% mutate(., Vpos = case_when(Vfin_position == "C" |  
                                     Vfin_position == "I" ~ "CI",  
                                     Vfin_position == "M" ~ "M",  
                                     Vfin_position == "F" ~ "F",  
                                     Vfin_position == "dna" ~ "dna")  
  ) -> WOS.ci
```

```
## Patterns without S: WOnoS.df
```

```
WOnoS.df <- bind_rows(DO, V)
```

```
WOnoS.df %>% mutate(., Vpos = case_when(Vfin_position == "C" |  
                                     Vfin_position == "I" ~ "CI",  
                                     Vfin_position == "M" ~ "M",  
                                     Vfin_position == "F" ~ "F",  
                                     Vfin_position == "dna" ~ "dna")  
  ) -> WOnoS.ci
```

4 H1: Distance between connective and finite verb

Provided canonical complementizers occupy clause-initial position, they should allow for a larger distance to the finite verb than elements that primarily modify the finite verb and tend to occupy a position adjacent to it. Thus, there should be an observable difference in terms of distance to the finite verb between more canonical complementizers (COMP) and less canonical ones (DIR).

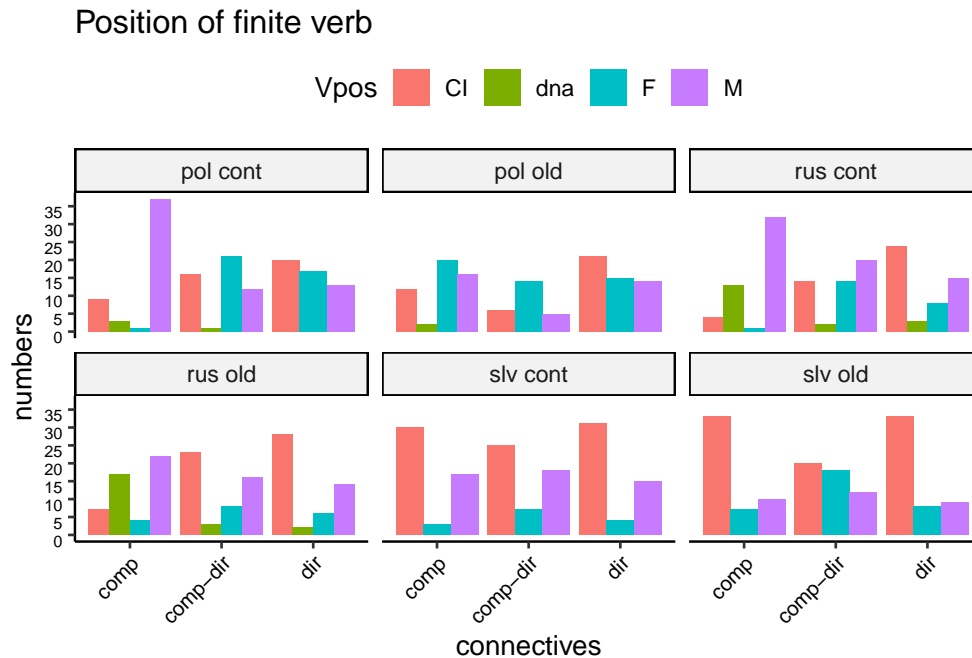


Figure 1: Position of finite verb, i.e. I(nitial)/C(litic preceded), M(edial), F(inal), for connectives across samples (dna = no finite Verb); in Russian and Polish, the finite verb tends to be placed later in the clause for COMP

Syllables between connective and finite verb across languages

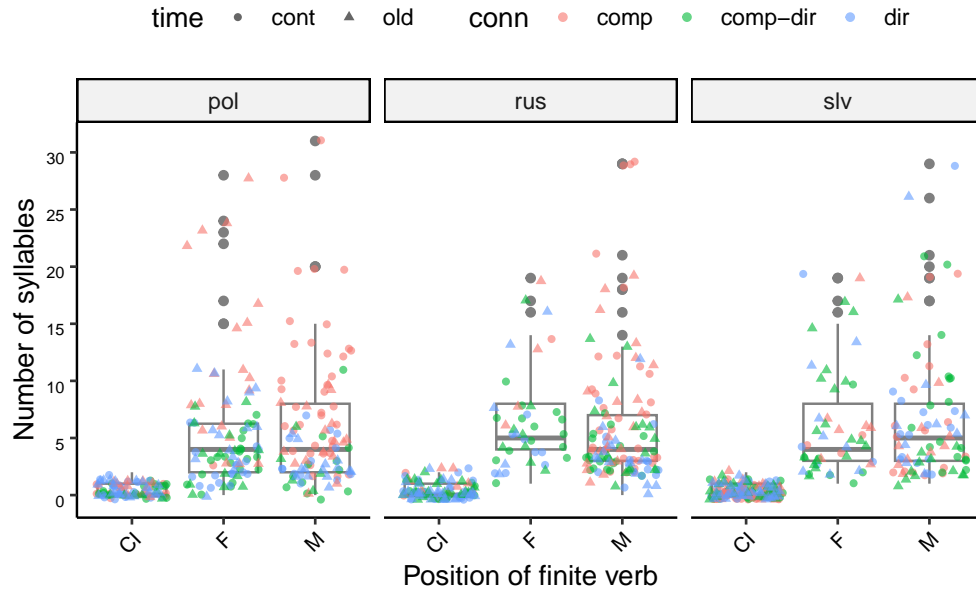


Figure 2: Syllables intervening between connective and finite verb considering the position of the finite verb across languages; more intervening syllables for M(edial) position of finite verb possible in all languages

Syllables intervening between connective and finite verb

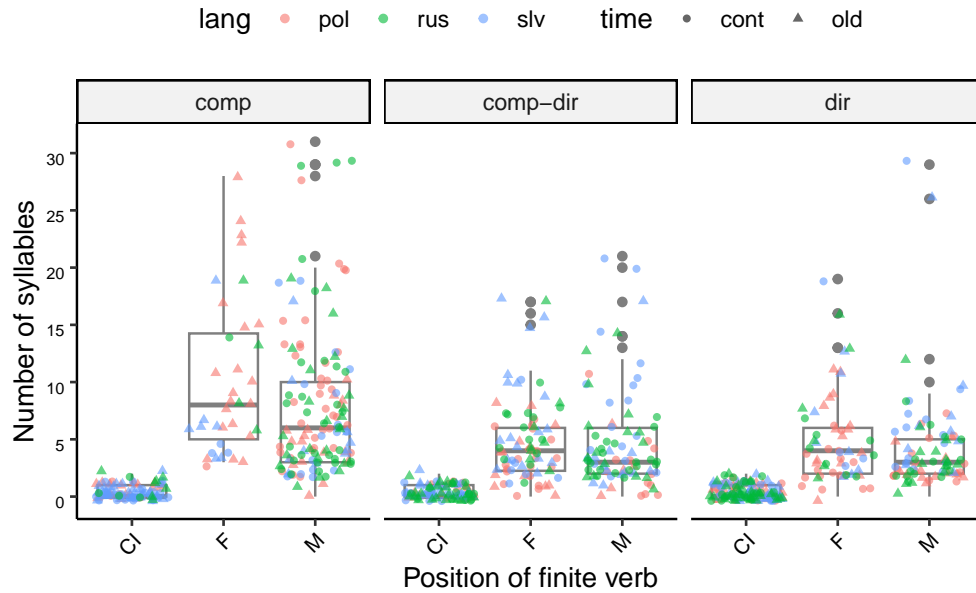


Figure 3: Syllables intervening between connective and finite verb considering the position of the finite verb across connectives; tendency towards more intervening syllables for COMP

Connectives compared to position of finite verb

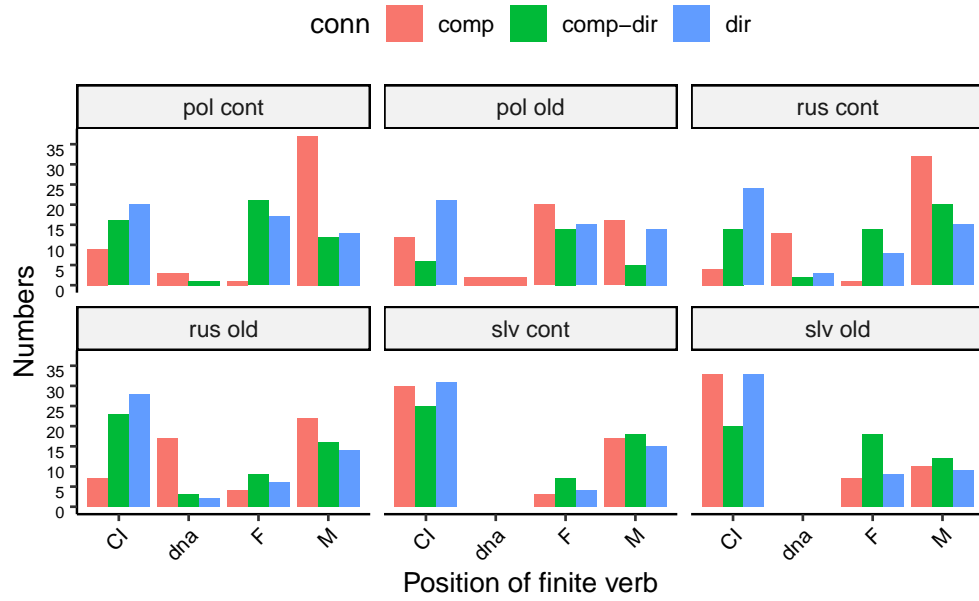


Figure 4: Position of finite verb for connectives across samples; connectives prefer I(nitial)/C(litic preceded) position of finite verb in Slovene

Syllables intervening between connective and finite verb

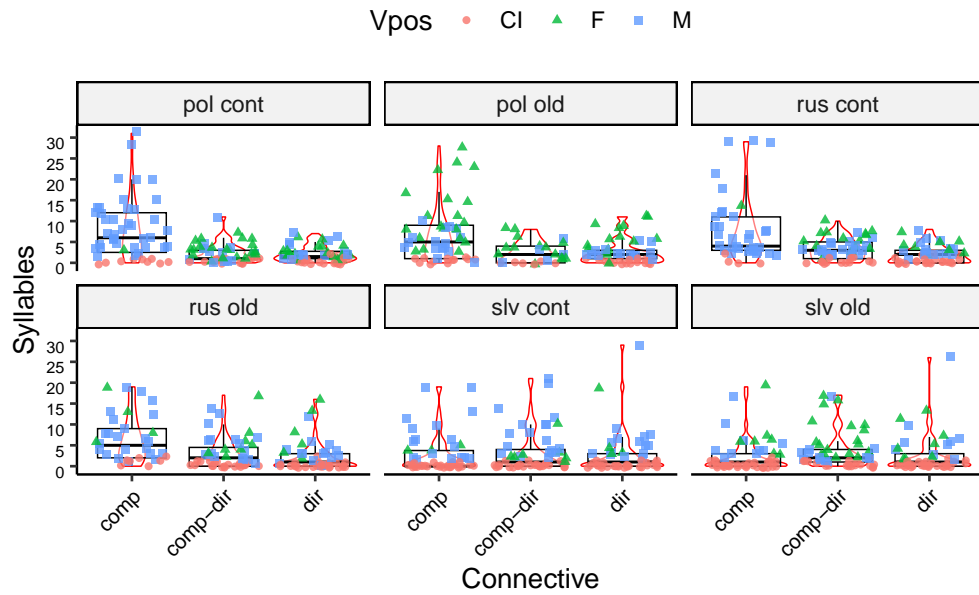
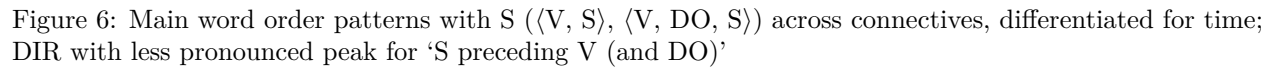


Figure 5: Syllables intervening between connective and position of finite verb for connectives across for samples; Slovene with most uniform patterns

If there is no difference in word order for main and subordinate clauses in Slavic, we should not find any difference concerning the position of elements in COMP- and DIR-clauses.



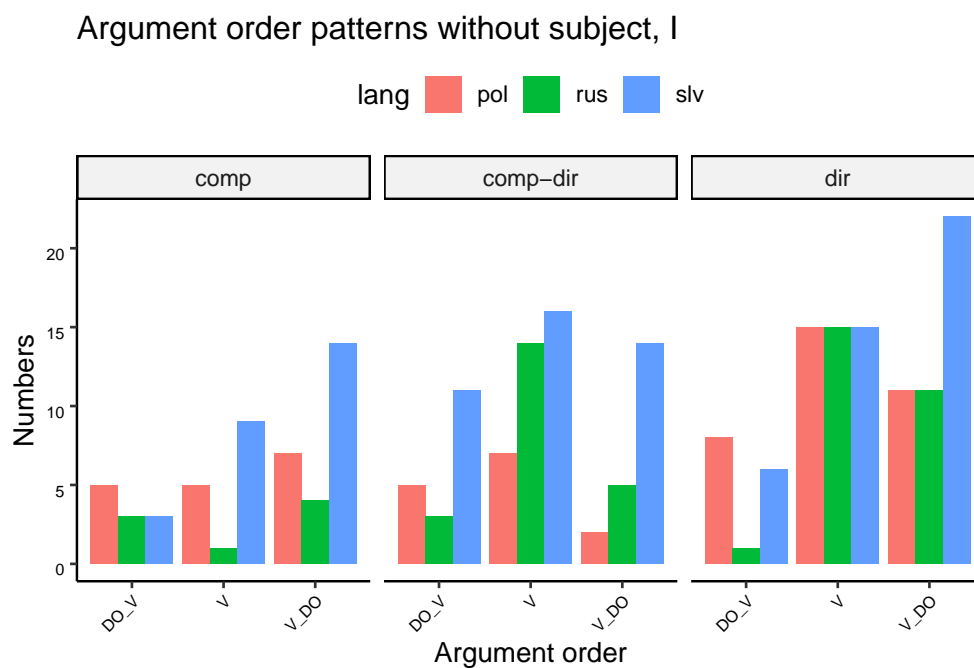


Figure 7: Main word order patterns without S ($\langle V, DO \rangle$, $\langle V \rangle$) across connectives, differentiated for language; slight preference for DIR to have 'V preceding DO'

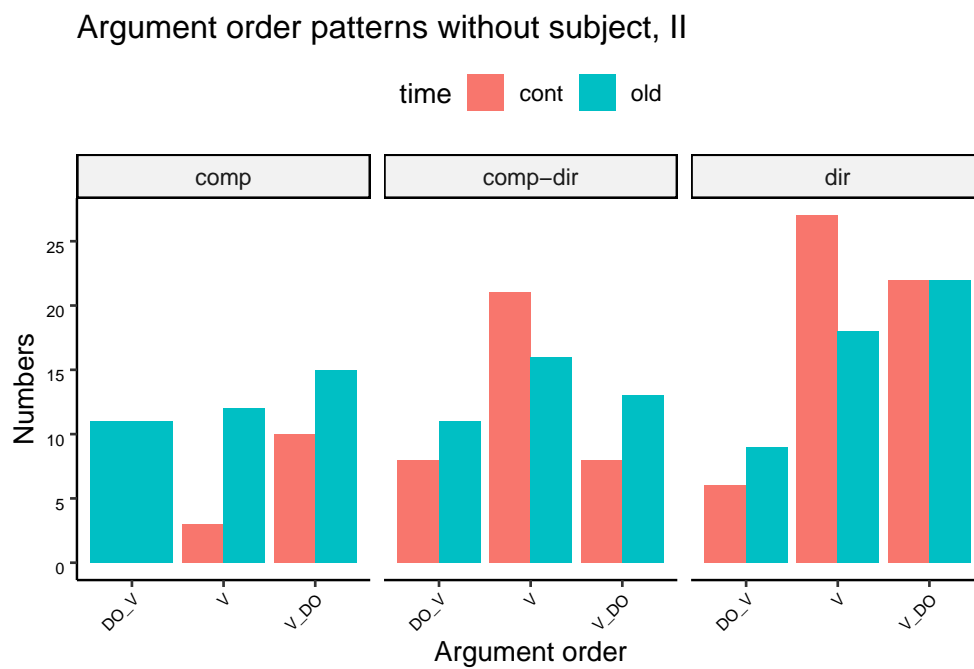


Figure 8: Main word order patterns without S ($\langle V, DO \rangle$, $\langle V \rangle$) across connectives, differentiated for time

6 H3: Position of finite verb

Clauses with adjacent COMP and DIR show more variation with respect to the position of the finite verb than either of them by itself, because their positions are influenced by two different principles (see H1).

Dispersion of distance between connective and finite verb

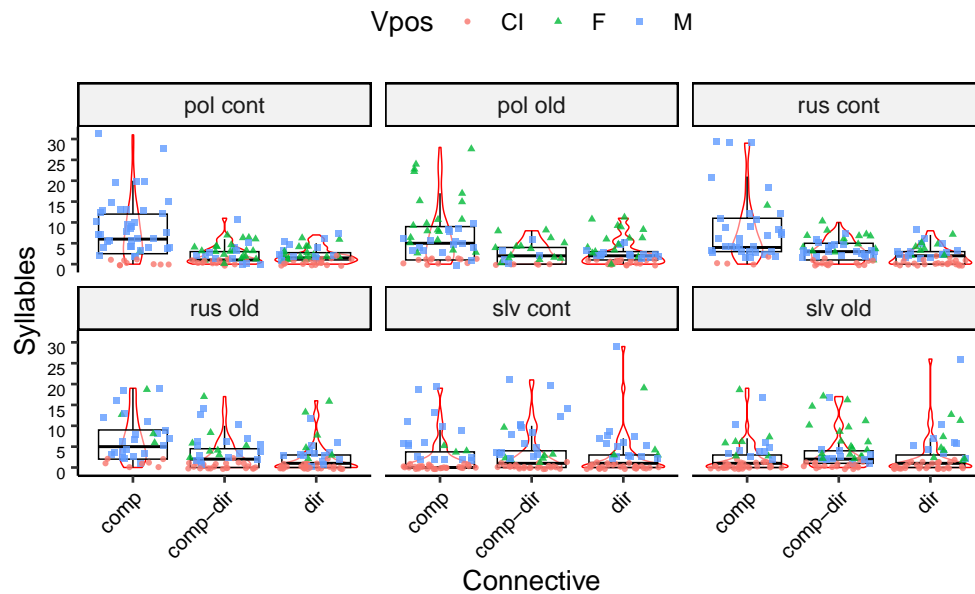


Figure 9: Syllables intervening between connective and position of finite verb for connectives across samples; greater dispersion for COMP than for DIR and COMP-DIR, except for Slovene

7 H4: Development of patterns

If patterns change over time, the patterns for DIR and COMP-DIR converge with the patterns for COMP (because they start behaving like canonical complementizers).

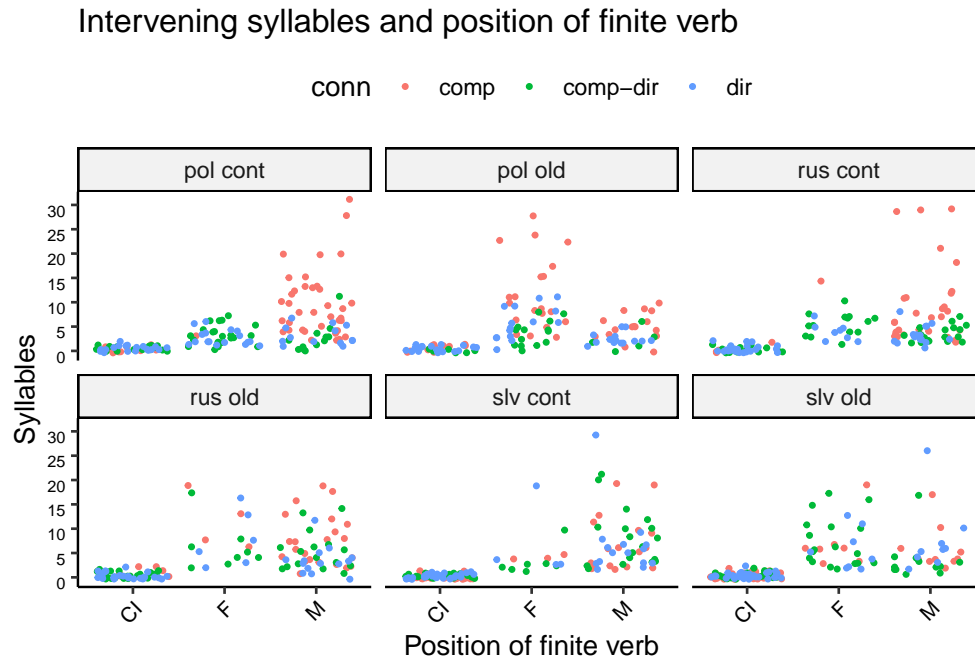


Figure 10: Syllables intervening between connective and finite verb for position of finite verb across samples; change from F(inal) to M(edial) position of finite verb for all languages, increase in number of syllables for Vpos=M in Polish and Russian in particular for COMP

8 H5: Frequency biases with CPTs

From a synchronic perspective, there are frequency biases in the occurrences of COMP, COMP-DIR and clause-initial DIR in the immediate vicinity of potential CTPs.

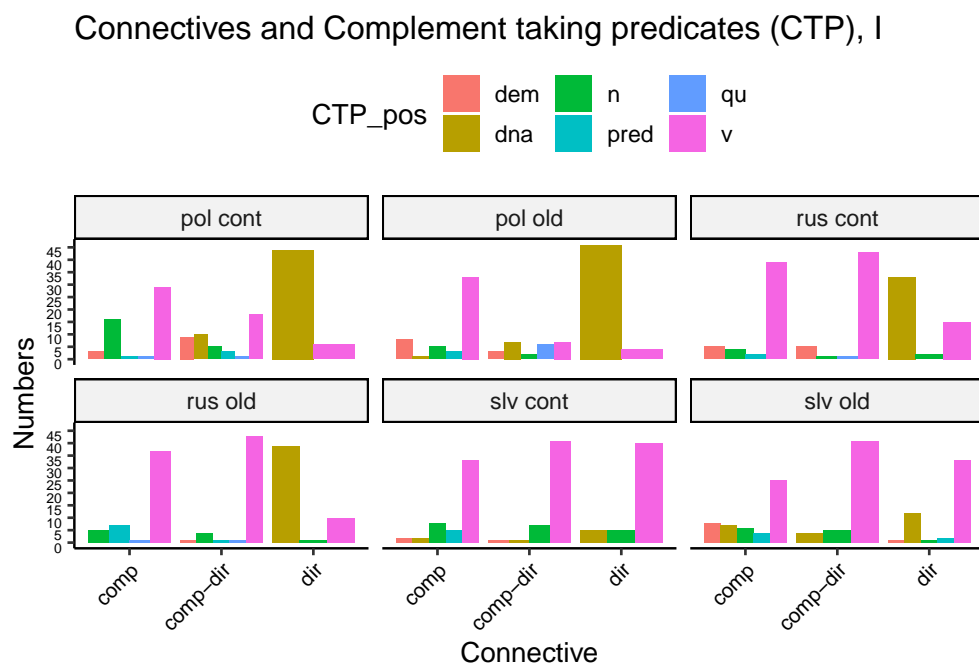


Figure 11: Part of speech of CTP for connectives across samples, part of speech = dna follows from absence of suitable CTP in surrounding context; DIR shows a tendency towards part of speech = dna

9 H6: Frequency shifts for CTPs

In diachronic comparison, there are frequency shifts for identical syntactic types of potential CTPs (regarding the same language).

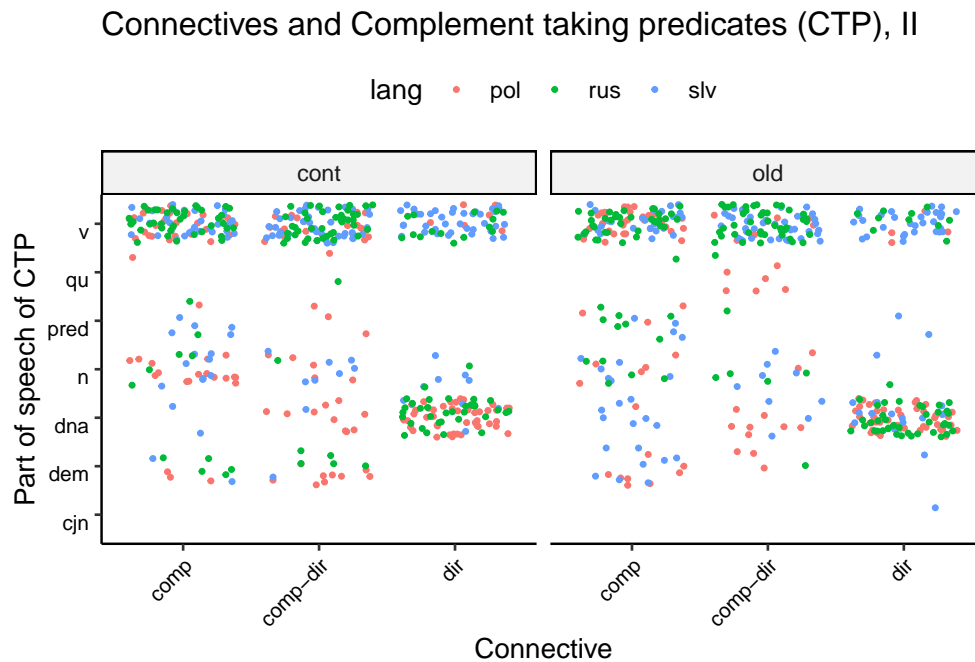


Figure 12: Part of speech of CTP for connectives and languages, across old and contemporary samples; single trends but hardly any patterns

10 References