Case study: case

Molly Rolf

September 2023

Contents

1 Intro

The repository for the third case study of the paper, case, contains (i) the two plots included in the paper, (ii) the two corresponding dataframes (adpo_props_facet.csv and adpo_props.csv), (iii) all the annotated texts (annotated_texts.csv), as well as data analysis and plotting code (this file).

The current file also details the corpus data used and how it has been manipulated by the authors.

2 Corpus data

The corpus data is from Annotated Corpus of Pre-Standardized Balkan Slavic Literature, which is available for download from https://gitlab.uzh.ch/uzh-slavic-corpora/pre-standardized-balkan-slavic-literature). Many of the corpus texts can be further explored through the online interface at https://www.punco.uzh.ch/. 19 texts from this corpus are included in the data in this paper, dated from c. 1340 to 1860. These are listed below with the authors' numbering, text title/detail and date:

24. Manasses Chronicle- Legend of Troy, Pověsti o izvěstovannyix veštei eže o kralexъ pričę. c.1340

31. Zograph Miscellany, žitïe î+ žíznь+ prpdbnyǫ+ matere+ nášę+ para_skevî. 1380s

32. Slovo Paraskevi, sˈty˝jeî prpdbniî paraskéviì. 1500s

29. Šimon Kožičić, Šimona Kožičića Zadranina biskupa Modruškoga: od žitiě rimskig' arxierěov' i cesarov'. 1531

20. Tixon.d. Žitie prěpodobnye matere našee Petky Trьnovskye. 1650-1670s

26. Tixon.d.damian, Рамеtь styx bezsrebrьnikь Kozma i Damianь. 1660s

35. NBKM 709, žítïe prpo_dobnïe mátere náši pétky trьnov'sky. 1690s

8. NBKM 328, Radi blaženoju Taïsïju. 1749-50

19. Temski r. Slovo o nakazanii dětei. 1764

9. NBKM 370, Istoria Slavěnobolgarskaa - Predislovie. 1784

16. PPS, Pop Punčov Sbornik. 1796

1. Berl.d., Slóvo styę prpdbnïę mtre nášīę Pét'kĭ. 1791-1803

15. Nedelnik1806, Skazanie o žitie prepodobnyę matere našeę Paraskevi Ternovskię. 1806

5. Krčovski 1814, Slovo vtoroe svętago proroka Daniila. 1814

11. NBMK 1064, Živeenitu i na sfetae prepodobnae maika našja Petka Tarnovskaja. 1820s

13. NBKM 1081, Homily of St. Prophet Daniel, Homily of a certain hermit. 1821

10. NBKM 728, Žitie prepodobnaę Paraskeva. 1820s-30s

23. Nedělnik 1856, Oktomvrię 14. za svętaę Paraskeva. 1856

17. Rai.d., Slovo ftoroe zaradi Gospodovata sadovie sfetago proroka Daniila. 1860

3 Data manipulation

The authors have made the following changes to the texts in the corpus as follows. Most importantly for the paper, each adpositional and nominal phrase was tagged with one of a number of function tags reflecting their grammatical role in the bigger phrase/clause (as opposed to morphological case marking). These are found in the func column of the annotated_texts.csv file. Each function tag is briefly described below:

Subj- subject. Tags what is often considered the syntactic or grammatical subject of a sentence. These can be intransitive or transitive subjects and are mostly agentive.

Dobj- direct object. Tags what is often considered the syntactic or grammatical direct object of a sentence. Less transitive 'objects' are marked with the goal tag.

Poss- possessive. Tags the possessor, whether this is a genitively marked noun or a possessive adjective, and includes partitive relations.

Recip- recipient. Tags the indirect object of the verb 'to give'.

Goal- goal. Tags less transitive 'objects' of less transitive verbs e.g. 'pray to', 'worship', 'follow', 'command'. Instances concerning physical movement or certain abstract movements to a location are tagged with the loc tag.

Loc- locative. Tags spatial relations and movements to and from locations e.g. 'in the world', 'into the desert' as well as locative-temporal relations such as 'on Tuesday'.

Abl- ablative. Tags instances of abstract movement away from a point in time or space as well as causative relations. Movement away from locations, especially involving physical movement, is tagged with the loc tag.

Ins- instrumental. Tags the stereotypical use of the instrumental morphological case i.e. that of an instrument or method by which an action is completed, such as 'they paganized their baptism with accursed divination and magic'. These instances are mostly inanimate, with animate instances tagged by an agentive tag (not included in the plots or analysis of this paper).

Ass- associative. Tags phrases of accompaniment or co-presence e.g. 'And he sent St. Cosmas and Damian to their home with great honors.'

Com- comitative. Tags the stereotypical use of the comitative case but is restricted to agentive instances- in other words, instances where both parties are active in the action e.g. 'I do not dare to do so, to go with you'. Inanimate or non-agentive instances are tagged with the ass tag.

In addition to the function tags, each nominal phrase was tagged with respect to whether it is marked with an adposition or not (y or n) in the `nominal_phrase_adpo_marked` column. Further annotation additions made by the authors are the `func_detail`, `further_func`, `case_of_nom`, further_caseandprep_marked' columns, none of which are used in the plots of this paper. Finally, the text names were added according to the authors' numbering and the text date (given in the corpus metadata). The annotated_texts file therefore includes all the changes made to the corpus by the authors. Several columns from the original corpus texts are not included in this file as they are not used by the authors.

The proportion of adpositional realization of each function listed above was then calculated using the `nominal_phrase_adpo_marked` and `func` columns to give the proportion (`prop`) in the `adpo_prop` and `adpo_prop_facet` dataframes. A number of other functions are not included in these dataframes for reasons of data size. The agentive tag mentioned above is not often used, for instance. Instances where the function is unclear (`unk`) are also disregarded. The dates of the texts in these two dataframes are given as single years; in the case of time periods, a midpoint is used.

The adpo_props_facet.csv file and corresponding plot includes the proportions for all the functions listed above whilst the adpo_props.csv and corresponding plot includes only `com`, `ins`, `goal`, `recip` and `poss`.

4   Data analysis and plots

```r
library(ggplot2)
library(tidyverse)
library(dplyr)
library(ggsci)

# read data
annotated_texts <- read.csv("annotated_texts.csv")

# grouping funcs from the annotated_texts file
annotated_texts$func[annotated_texts$func == "path"] <- "loc"
```

```r
annotated_texts$func[annotated_texts$func == "pathplace"] <- "loc"
annotated_texts$func[annotated_texts$func == "temp"] <- "loc"
annotated_texts$func[annotated_texts$func == "path:goal"] <- "loc"
annotated_texts$func[annotated_texts$func == "path:source"] <- "loc"
annotated_texts$func[annotated_texts$func == "caus"] <- "abl"

# deleting funcs from the annotated_texts file
annotated_texts <- annotated_texts[!(annotated_texts$func == "about"), ]
annotated_texts <- annotated_texts[!(annotated_texts$func == "accord"), ]
annotated_texts <- annotated_texts[!(annotated_texts$func == "agent"), ]
annotated_texts <- annotated_texts[!(annotated_texts$func == "ben"), ]
annotated_texts <- annotated_texts[!(annotated_texts$func == "comp"), ]
annotated_texts <- annotated_texts[!(annotated_texts$func == "instead"), ]
annotated_texts <- annotated_texts[!(annotated_texts$func == "purp"), ]
annotated_texts <- annotated_texts[!(annotated_texts$func == "simul"), ]
annotated_texts <- annotated_texts[!(annotated_texts$func == "unk"), ]
annotated_texts <- annotated_texts[!(annotated_texts$func == "unm"), ]
annotated_texts <- annotated_texts[!(annotated_texts$func == "until"), ]
annotated_texts <- annotated_texts[!(annotated_texts$func == "voc"), ]
annotated_texts <- annotated_texts[!(annotated_texts$func == "without"), ]
annotated_texts <- annotated_texts[!(annotated_texts$func == "voc"), ]


# calculate proportion of adpositional realizatin for each remaining function
adpo_props_facet <- annotated_texts %>%
    group_by(textname, date, func) %>%
    summarise(prop = mean(nominal_phrase_adpo_marked == "y", na.rm = TRUE))
# delete NA row
adpo_props_facet <- na.omit(adpo_props_facet)

# delete further functions for second plot
adpo_props <- adpo_props_facet[!(adpo_props_facet$func == "subj"), ]
adpo_props <- adpo_props[!(adpo_props$func == "dobj"), ]
adpo_props <- adpo_props[!(adpo_props$func == "abl"), ]
adpo_props <- adpo_props[!(adpo_props$func == "ass"), ]
adpo_props <- adpo_props[!(adpo_props$func == "loc"), ]
```

```r
# write data
write.csv(adpo_props, "adpo_props.csv", row.names = FALSE)
write.csv(adpo_props_facet, "adpo_props_facet.csv", row.names = FALSE)


# factorize functions
adpo_props$func <- factor(adpo_props$func)
adpo_props_facet$func <- factor(adpo_props_facet$func)

# rename levels
levels(adpo_props$func) <- c("comitative", "goal", "instrumental", "possessive",
    "recipient")
levels(adpo_props_facet$func) <- c("ablative", "associative", "comitative", "direct object",
    "goal", "instrumental", "locative", "possessive", "recipient", "subject")

# change order of funcs
adpo_props$func <- factor(adpo_props$func, levels = c("comitative", "instrumental",
    "goal", "recipient", "possessive"))
adpo_props_facet$func <- factor(adpo_props_facet$func, levels = c("comitative", "associative",
    "instrumental", "ablative", "locative", "goal", "recipient", "possessive", "direct object",
    "subject"))

# adpositional proportions facet plot (all functions)
d <- ggplot(adpo_props_facet, aes(x = date, y = prop)) + geom_point(color = "#3d62a6",
    size = 1) + geom_smooth(fullrange = FALSE, se = TRUE, span = 1.5, color = "black",
    size = 0.8) + facet_wrap(. ~ func, nrow = 2) + theme_bw() + theme(axis.text = element_text(color = "black")) +
    ylim(0, 1) + scale_y_continuous(expand = c(0, 0), limits = c(-1, 2)) + coord_cartesian(ylim = c(-0.03,
    1.03)) + ylab("Proportion of adpositional realization") + xlab("Time") + theme(axis.text.x = element_text(angle = 60,
    hjust = 1), strip.text = element_text(size = 10), strip.background = element_blank())
#+ geom_point(aes(size=no_of_NPs))

print(d)
```
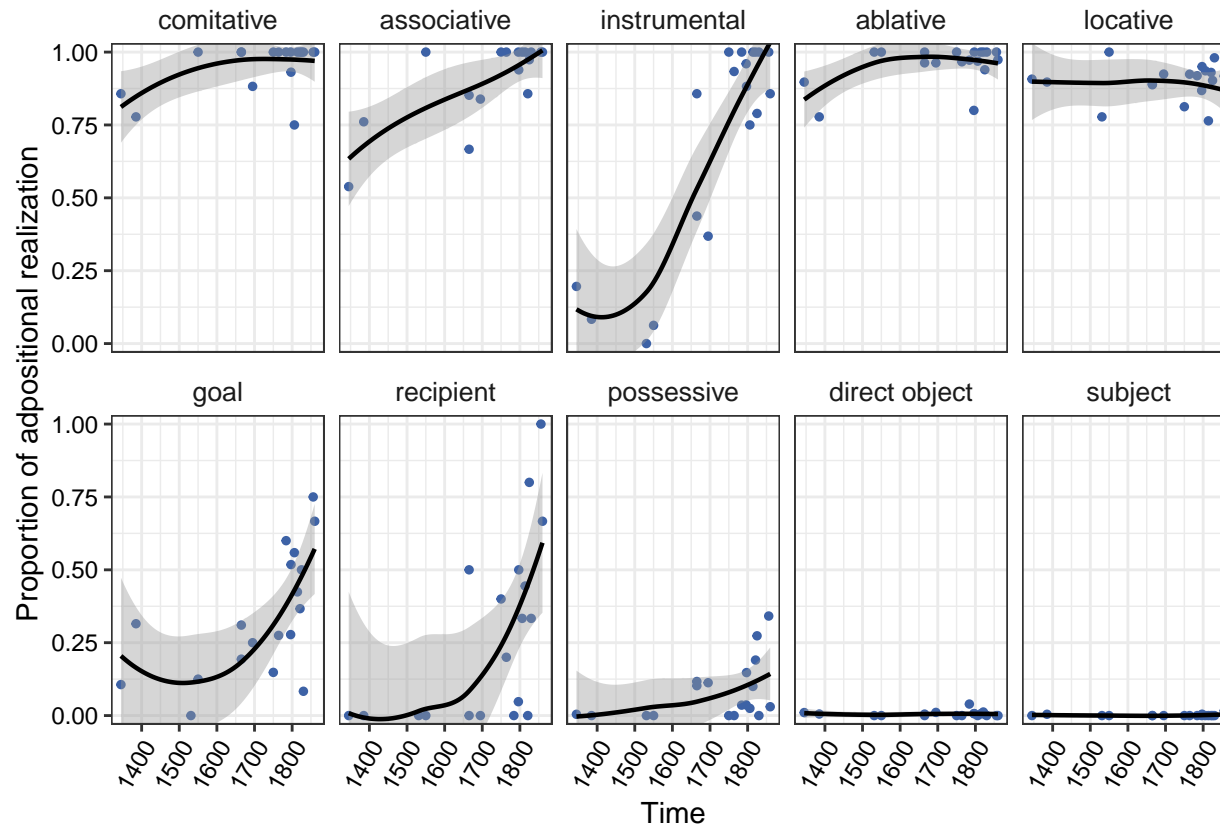
```
ggsave("Figure4.pdf", d, height = 4, width = 6.5)


# adpositional proportions second plot (com, ins, goal, recip and poss)
p <- ggplot(adpo_props, aes(x = date, y = prop, color = func, linetype = func, group = func)) +
    geom_smooth(se = FALSE, span = 1.5) + labs(y = "Proportion of adpositional realization",
    x = "Time") + theme_bw() + theme(axis.text = element_text(color = "black")) +
    ylim(0, 1) + scale_color_npg() + labs(linetype = "Function", color = "Function")

print(p)
```
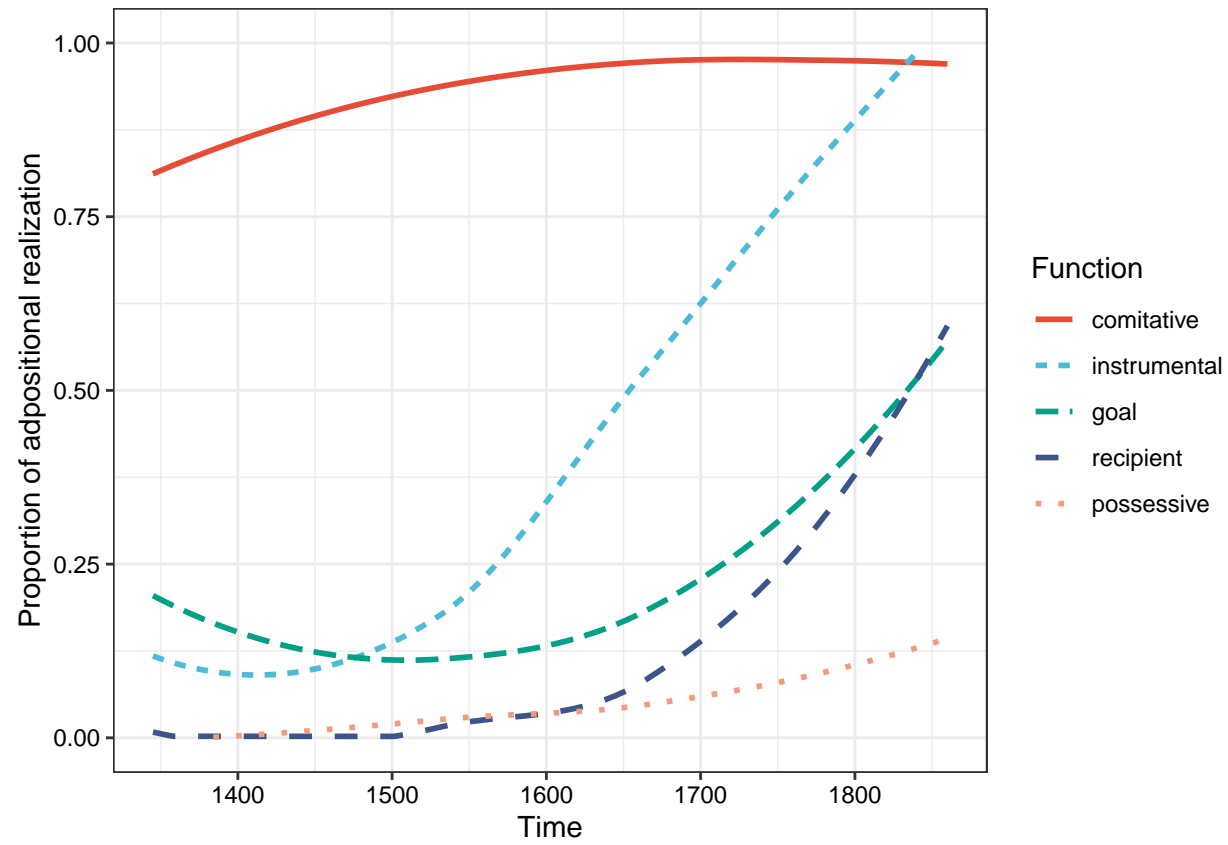
```
ggsave("Figure5.pdf", p, height = 3.5, width = 5)
```