

# Case study: number

## The Development of Syntagmatic Redundancy in ME

Raquel Montero

September 2023

### Contents

<b>1</b>	<b>Introduction</b>	<b>1</b>
<b>2</b>	<b>Data Cleaning</b>	<b>2</b>
<b>3</b>	<b>Data Visualization</b>	<b>6</b>
3.1	Map . . . . .	6
3.2	Barplots . . . . .	13
<b>4</b>	<b>Statistical Analysis</b>	<b>19</b>
4.1	Linear Regression Model (lm) . . . . .	19
4.2	Test of residuals (DHARMa) . . . . .	20

### 1 Introduction

This file shows the code used to analyse the data in Section 3 of the paper “Sociolinguistic typology meets historical corpus linguistics”; submitted to the Transactions of the Philological Society. Information on how the data was searched, annotated etc. can be found in the Github repository. The following is a guide to the documents that the repository contains:

- Annotation Guidelines: AnnotationGuidelinesSyntRedME.pdf
- CorpusSearch: CorpusSearchSyntRedME\_Adjectives.c//CorpusSearchSyntRedME\_Quantifiers.c
- Geographical Information about the texts: GeographicsSyntRedME.pdf
- Final data: SyntRedME\_Adjectives.cod.ooo//SyntRedME\_Quantifiers.cod.ooo

## 2 Data Cleaning

First, we need to load the packages that will be used:

```
library(gdata, warn.conflicts = FALSE) # for drop-levels
library(plyr, warn.conflicts = FALSE)
library(dplyr, warn.conflicts = FALSE)
library(ggplot2, warn.conflicts = FALSE)
library(ggrepel, warn.conflicts = FALSE)
library(splines, warn.conflicts = FALSE)
library(MASS, warn.conflicts = FALSE)
library(reshape2, warn.conflicts = FALSE)
library(maps, warn.conflicts = FALSE)
library(viridisLite, warn.conflicts = FALSE)
library(viridis, warn.conflicts = FALSE) # colors of the map
library(DHARMA, warn.conflicts = FALSE) # for residuals
library(colorspace)
suppressPackageStartupMessages(library(sjPlot)) # to change the font
```

Then, we import the annotated data, and prepare it for visualization (e.g. eliminating variables that are not needed, converting to numeric the scores, etc.):

```
## ADJECTIVES DATA:
```

```
# Importing the data:
```

```
data_agreementadj <- read.delim("Data/SyntRedME_Adjectives.cod.ooo.ooo", header=F, sep=":")
```

```
# Naming the columns:
```

```
colnames(data_agreementadj) <- c("position",
                                "agreement",
                                "Text",
                                "Year",
                                "Dialect",
                                "numsyll",
                                "animate")
```

```
# Eliminate variables not used:
```

```
ex.dataadj <- subset(data_agreementadj,
                    position != "z" &
```

```

position != "multiple"& # Eliminating cases with multiple adjectives
position != "Conjoined" & # Eliminating cases of conjoined adjectives
agreement!= "z" &
agreement!= "french-agree" & # Eliminating cases with French morphology
agreement != "adverbial-ly" & # Eliminating adjectives formed with -ly
agreement != "e-ending" & # Eliminate adjectives with root ends in -e
agreement != "uninflected" & # Eliminate adjectives uninflected in OE
Text != "z" &
Year != "z" &
Dialect != "z" &
numsyll == "monosyllabic" # Only monosyllabic adjectives
)

```

#### ## QUANTIFIERS DATA:

*# Importing the data:*

```
data_agreementquant <- read.delim("Data/SyntRedME_Quantifiers.cod.ooo.ooo", header=F, sep=":")
```

*# Naming the columns:*

```
colnames(data_agreementquant) <- c("quant",
                                   "agreement",
                                   "Text",
                                   "Year",
                                   "Dialect",
                                   "numsyll")
```

*# Eliminate variables not used:*

```
ex.dataquant <- subset(data_agreementquant,
                       quant != "z" &
                       agreement != "z" &
                       Text != "z" &
                       Year != "z" &
                       Dialect != "z" &
                       numsyll == "monosyllabic"
                       )
```

```

# droplevels: to drop unused factor levels.
ex.dataadj <- droplevels(ex.dataadj)
ex.dataquant <- droplevels(ex.dataquant)

# years stored as numbers:
ex.dataadj$Year <- as.numeric(as.character(ex.dataadj$Year))
ex.dataquant$Year <- as.numeric(as.character(ex.dataquant$Year))

# create a new column changing the agreement type into a binary variable: (agreement = 1, no agreement = 0)
ex.dataadj$agreement_new <- ifelse(ex.dataadj$agreement == "agreement", 1, 0)
ex.dataadj$agreement_new <- as.numeric(as.character(ex.dataadj$agreement_new)) # change into numeric value
ex.dataquant$agreement_new <- ifelse(ex.dataquant$agreement == "agreement", 1, 0)
ex.dataquant$agreement_new <- as.numeric(as.character(ex.dataquant$agreement_new)) #change into numeric value

```

We then have a look at how much data there is for our data analysis:

```

table <- table(ex.dataadj$Dialect)
table

##
##      EM Northern Southern      WM
##      528         78      276    300

```

```

table1 <- table(ex.dataquant$Dialect)
table1

##
##      EM Northern Southern      WM
##     1535         272      599    778

```

Calculate the mean agreement per text

```

plot.data2 <- ddply(ex.dataadj, .(Text, Dialect, Year), summarize, whet = mean(as.numeric(as.character(agreement_new))),
  na.rm = T), n = sum(!is.na(as.numeric(as.character(agreement_new)))))
plot.data2

```

```

##      Text  Dialect Year      whet  n
## 1      Aelred      EM 1325 0.77777778 18
## 2      Ancrene      WM 1230 0.94736842 57
## 3    Astrolabe      EM 1390 0.83333333  6
## 4    Ayenbyte Southern 1340 0.96969697 66

```

## 5	Benrul	Northern	1425	0.87500000	8
## 6	Boethious	EM	1380	1.00000000	12
## 7	Book-vices	EM	1400	0.87500000	8
## 8	Brut	WM	1400	0.84210526	38
## 9	Capgrave-chnicle	EM	1460	0.60975610	41
## 10	Chronicles	EM	1150	1.00000000	7
## 11	Cloud	EM	1400	1.00000000	4
## 12	Edmund	WM	1390	1.00000000	2
## 13	Edmund-m4	EM	1450	1.00000000	2
## 14	Edthor	Northern	1350	0.80000000	5
## 15	Equatorie	EM	1390	1.00000000	2
## 16	Gaytry	Northern	1440	0.75000000	8
## 17	Gregory	Southern	1475	0.91428571	35
## 18	Hilton	EM	1390	1.00000000	3
## 19	Horses	Southern	1450	0.80000000	5
## 20	Innocencium	Southern	1490	0.77777778	9
## 21	Katherine-group	WM	1225	0.93333333	15
## 22	Kempe	EM	1450	0.32394366	71
## 23	Kentish-homilies	Southern	1150	1.00000000	3
## 24	Kentish-sermons	Southern	1275	1.00000000	3
## 25	Lambeth	WM	1225	0.96774194	31
## 26	Mandeville	EM	1400	0.85393258	89
## 27	medicinis	Northern	1440	1.00000000	7
## 28	Melibee	EM	1390	0.94444444	18
## 29	Mirk	WM	1415	0.09589041	73
## 30	Morte-a	WM	1470	0.55405405	74
## 31	New-testament	EM	1380	0.66666667	6
## 32	Old-testament	EM	1380	1.00000000	6
## 33	Parson	EM	1390	0.92307692	26
## 34	Polychronicon	Southern	1380	0.78787879	33
## 35	Psalter	EM	1350	0.56666667	30
## 36	Purvey	Southern	1380	0.87179487	117
## 37	Reynard	EM	1480	0.84615385	13
## 38	Reynes	EM	1480	0.63636364	11
## 39	Richard-rolle	Northern	1350	0.78000000	50
## 40	Royal	Southern	1425	0.20000000	5
## 41	Siege-j	WM	1490	0.20000000	10
## 42	Trinity	EM	1225	0.93478261	46
## 43	Vices	EM	1225	1.00000000	24

```
## 44      wycliffite      EM 1400 0.90588235 85
```

```
plot.data4 <- ddply(ex.dataquant, .(Text, Dialect, Year), summarize, whet = mean(as.numeric(as.character(agreement_new))),  
  na.rm = T), n = sum(!is.na(as.numeric(as.character(agreement_new)))))  
head(plot.data4)
```

```
##      Text  Dialect Year      whet  n  
## 1   Aelred      EM 1325 0.9000000 30  
## 2 Aelred-m4      EM 1450 0.9090909 22  
## 3  Ancrene      WM 1230 0.9000000 120  
## 4 Astrolabe      EM 1390 0.8947368 19  
## 5  Ayenbyte Southern 1340 0.9467456 169  
## 6   Benrul Northern 1425 0.7222222 54
```

Merging the two data sets:

```
## First, merge the two data frames:
```

```
total <- merge(plot.data2, plot.data4, by = c("Text", "Dialect", "Year"))
```

```
colnames(total) <- c("Text", "Dialect", "Year", "adjectives", "nad", "quantifier",  
  "nq")
```

```
## New column with the differences in the agreement proportion:
```

```
total$diff <- total$adjectives - total$quantifier
```

## 3 Data Visualization

### 3.1 Map

First we need to add a column with the coordinates for each text:

```
# Adding column based on new Coordinates:
```

```
#Adding the Latitude:
```

```
total4 <- total %>%  
  mutate(Longitude = case_when(  
    (Text== "medicinis") ~ "-0.56880000",  
    (Text== "Norwich") ~ "-0.20312085",  
    (Text== "Horses" ) ~ "-1.0318673",
```

```

(Text=="Richard-rolle") ~ "-2.0014878",
(Text=="Innocencium") ~ "-1.4743217",
(Text=="Edthor") ~ "-2.2412109",
(Text=="Mandeville") ~ "0.33441843",
(Text=="Edmund-m4") ~ "0.0000000",
(Text=="Gaytry") ~ "-2.2412109",
(Text=="Capgrave-chnicle") ~ "1.0000000",
(Text=="Capgrave-sermon") ~ "0.41325972",
(Text=="Morte-a") ~ "-1.5536906",
(Text=="Brut") ~ "-2.9340404",
(Text=="Kentish-homilies") ~ "-1.8031714",
(Text=="Royal") ~ "-0.80603102",
(Text=="Polychronicon") ~ "-2.4498470",
(Text=="Benrul") ~ "-1.0836749",
(Text=="Chronicles") ~ "-0.24573507",
(Text=="Gregory") ~ "-0.60829082",
(Text=="Cloud") ~ "-1.3818890",
(Text=="Mirk") ~ "-1.8076144",
(Text=="Lambeth") ~ "-2.4116106",
(Text=="Ancrene") ~ "-2.7064436",
(Text=="Siege-j") ~ "-1.6660928",
(Text=="Ayenbyte") ~ "1.0814800",
(Text=="Edmund") ~ "-2.1569555",
(Text=="Katherine-group") ~ "-2.7064436",
(Text=="Boethious") ~ "-0.12772404",
(Text=="Old-testament") ~ "-0.89464405",
(Text=="Vices") ~ "0.12773182",
(Text=="Equatorie") ~ "-0.67179476",
(Text=="Hilton") ~ "-0.22077060",
(Text=="Kentish-sermons") ~ "0.62063882",
(Text=="Parson") ~ "-0.12772404",
(Text=="Trinity") ~ "-1.6118882",
(Text=="Melibee") ~ "-0.12772404",
(Text=="Book-vices") ~ "0.11872816",
(Text=="Astrolabe") ~ "-1.0318673",
(Text=="wycliffite") ~ "-1.3150252",
(Text=="Purvey") ~ "-0.12942373",
(Text=="Reynard") ~ "0.0000000",

```

```

(Text=="Aelred") ~ "-2.3237144",
(Text=="Aelred-m4") ~ "0.0000000",
(Text=="Reynes") ~ "1.5482236",
(Text=="Psalter") ~ "0.0039963911",
(Text=="New-testament") ~ "-0.27629328",
(Text=="Kempe") ~ "1.0000000",

))

```

*#Adding the Longitude:*

```

total5 <- total4 %>%
  mutate(Latitude = case_when(
    (Text== "medicinis") ~ "53.065628",
    (Text== "Horses" ) ~ "51.453486",
    (Text== "Norwich") ~ "-53.182303",
    (Text=="Richard-rolle") ~ "53.955093",
    (Text=="Innocencium") ~ "51.068150",
    (Text=="Edthor") ~ "54.690768",
    (Text=="Mandeville") ~ "52.376964",
    (Text=="Edmund-m4") ~ "52.505077",
    (Text=="Gaytry") ~ "54.690768",
    (Text=="Capgrave-chnicle") ~ "52.666667",
    (Text=="Capgrave-sermon") ~ "52.744048",
    (Text=="Morte-a") ~ "52.321306",
    (Text=="Brut") ~ "52.018714",
    (Text=="Kentish-homilies") ~ "50.892692",
    (Text=="Royal") ~ "51.485871",
    (Text=="Polychronicon") ~ "51.688892",
    (Text=="Benrul") ~ "53.962009",
    (Text=="Chronicles") ~ "52.566821",
    (Text=="Gregory") ~ "51.357801",
    (Text=="Cloud") ~ "52.560221",
    (Text=="Mirk") ~ "53.029131",
    (Text=="Lambeth") ~ "52.255453",
    (Text=="Ancrene") ~ "52.370931",
    (Text=="Siege-j") ~ "52.039927",

```



```

(Text=="Ayenbyte") ~ "51.280552",
(Text=="Edmund") ~ "52.210949",
(Text=="Katherine-group") ~ "52.370931",
(Text=="Boethious") ~ "51.507407",
(Text=="Old-testament") ~ "51.783499",
(Text=="Vices") ~ "51.742427",
(Text=="Equatorie") ~ "52.796561",
(Text=="Hilton") ~ "52.161753",
(Text=="Kentish-sermons") ~ "51.246812",
(Text=="Parson") ~ "51.507407",
(Text=="Trinity") ~ "51.563236",
(Text=="Melibee") ~ "51.507407",
(Text=="Book-vices") ~ "51.859491",
(Text=="Astrolabe") ~ "51.507407",
(Text=="wycliffite") ~ "53.063272",
(Text=="Purvey") ~ "52.178413",
(Text=="Reynard") ~ "52.505077",
(Text=="Aelred") ~ "52.255726",
(Text=="Aelred-m4") ~ "52.505077",
(Text=="Reynes") ~ "52.632148",
(Text=="Psalter") ~ "51.564752",
(Text=="New-testament") ~ "52.162668",
(Text=="Kempe") ~ "52.666667",
))

```

*## The coordinates are stored as characters but we need the numbers:*

```

total5$Latitude <- as.numeric(as.character(total5$Latitude))
total5$Longitude <- as.numeric(as.character(total5$Longitude))

```

Create the map:

*## Creating the map:*

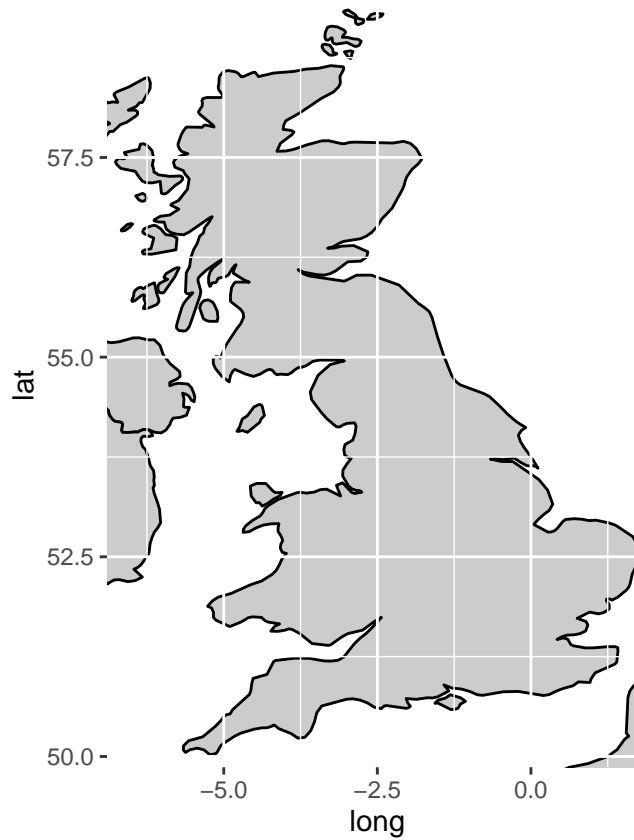
```

worldmap = map_data("world")
knitr::kable(head(worldmap, 20))

```

	long	lat	group	order	region	subregion
1	-69.89912	12.45200	1	1	Aruba	NA
2	-69.89571	12.42300	1	2	Aruba	NA
3	-69.94219	12.43853	1	3	Aruba	NA
4	-70.00415	12.50049	1	4	Aruba	NA
5	-70.06612	12.54697	1	5	Aruba	NA
6	-70.05088	12.59707	1	6	Aruba	NA
7	-70.03511	12.61411	1	7	Aruba	NA
8	-69.97314	12.56763	1	8	Aruba	NA
9	-69.91181	12.48047	1	9	Aruba	NA
10	-69.89912	12.45200	1	10	Aruba	NA
12	74.89131	37.23164	2	12	Afghanistan	NA
13	74.84023	37.22505	2	13	Afghanistan	NA
14	74.76738	37.24917	2	14	Afghanistan	NA
15	74.73896	37.28564	2	15	Afghanistan	NA
16	74.72666	37.29072	2	16	Afghanistan	NA
17	74.66895	37.26670	2	17	Afghanistan	NA
18	74.55899	37.23662	2	18	Afghanistan	NA
19	74.37217	37.15771	2	19	Afghanistan	NA
20	74.37617	37.13735	2	20	Afghanistan	NA
21	74.49796	37.05722	2	21	Afghanistan	NA

```
ggplot() + geom_polygon(data = worldmap, aes(x = long, y = lat, group = group), fill = "gray80",
  color = "black") + coord_fixed(ratio = 1.3, xlim = c(-6.5, 1.5), ylim = c(50.3,
  59)) + theme(panel.background = element_rect(fill = NA), panel.ontop = TRUE)
```



Add the information to the map:

```
# windowsFonts('Garamond' = windowsFont('Garamond'))

set_theme(base = theme_bw(), title.size = 1.4, axis.title.size = 1.3, axis.textsize = 1.1,
  axis.textcolor = "black", axis.title.color = "black", legend.size = 1.2, legend.title.size = 1.4,
  geom.label.size = 3)

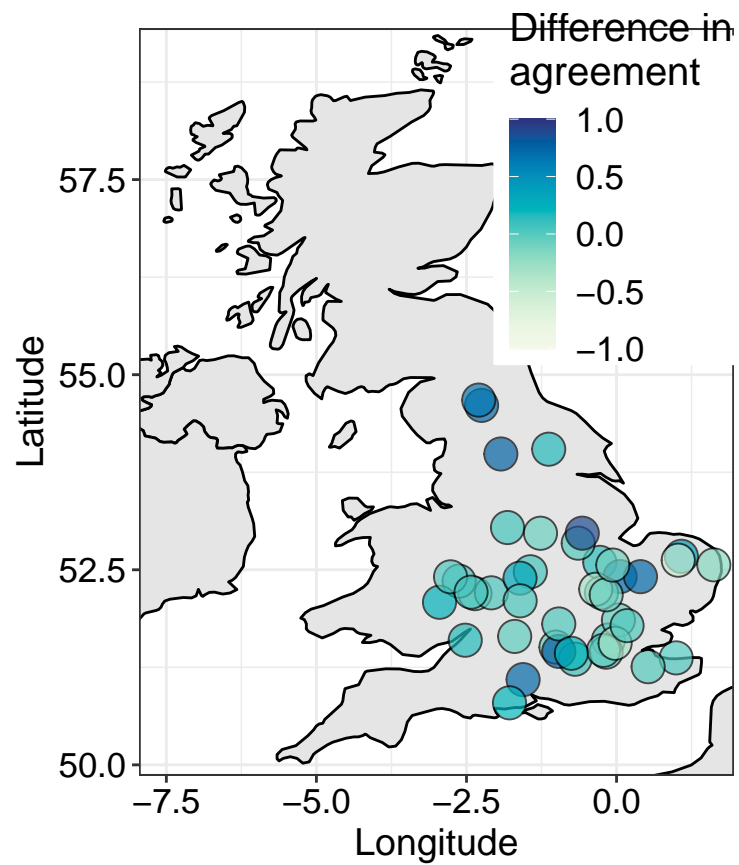
worldmap = map_data("world")
knitr::kable(head(worldmap, 20))
```

	long	lat	group	order	region	subregion
1	-69.89912	12.45200	1	1	Aruba	NA
2	-69.89571	12.42300	1	2	Aruba	NA
3	-69.94219	12.43853	1	3	Aruba	NA
4	-70.00415	12.50049	1	4	Aruba	NA
5	-70.06612	12.54697	1	5	Aruba	NA
6	-70.05088	12.59707	1	6	Aruba	NA
7	-70.03511	12.61411	1	7	Aruba	NA
8	-69.97314	12.56763	1	8	Aruba	NA
9	-69.91181	12.48047	1	9	Aruba	NA
10	-69.89912	12.45200	1	10	Aruba	NA
12	74.89131	37.23164	2	12	Afghanistan	NA
13	74.84023	37.22505	2	13	Afghanistan	NA
14	74.76738	37.24917	2	14	Afghanistan	NA
15	74.73896	37.28564	2	15	Afghanistan	NA
16	74.72666	37.29072	2	16	Afghanistan	NA
17	74.66895	37.26670	2	17	Afghanistan	NA
18	74.55899	37.23662	2	18	Afghanistan	NA
19	74.37217	37.15771	2	19	Afghanistan	NA
20	74.37617	37.13735	2	20	Afghanistan	NA
21	74.49796	37.05722	2	21	Afghanistan	NA

```
# set RNG seed so jitter is always in the same place
set.seed(3242)
```

```
mapengland2 <- ggplot() + geom_polygon(data = worldmap, aes(x = long, y = lat, group = group),
  fill = "grey90", color = "black") + coord_fixed(ratio = 1.3, xlim = c(-7.5, 1.5),
  ylim = c(50.3, 59)) + geom_point(data = total5, aes(x = Longitude, y = Latitude,
  fill = diff), alpha = I(0.7), size = 5.5, color = "black", pch = 21, position = position_jitter(h = 0.1,
  w = 0.1)) + scale_fill_continuous_sequential(palette = "GnBu", limits = c(-1,
  1)) + theme(legend.position = c(0.82, 0.77)) + labs(x = "Longitude", y = "Latitude",
  fill = expression(paste("Difference in \nagreement")))
```

```
mapengland2
```



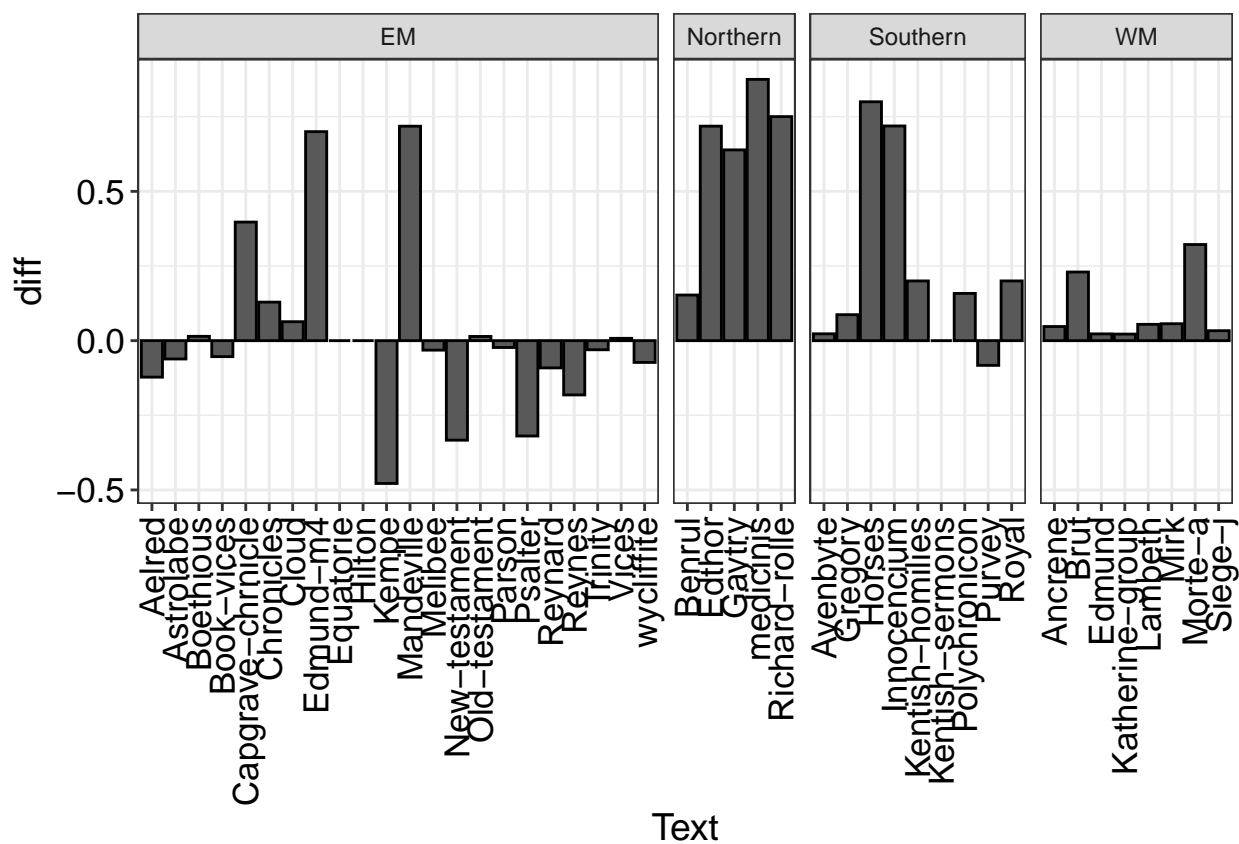
```
ggsave(mapengland2, file = "Figure1.pdf", width = 8, height = 8)
```

### 3.2 Barplots

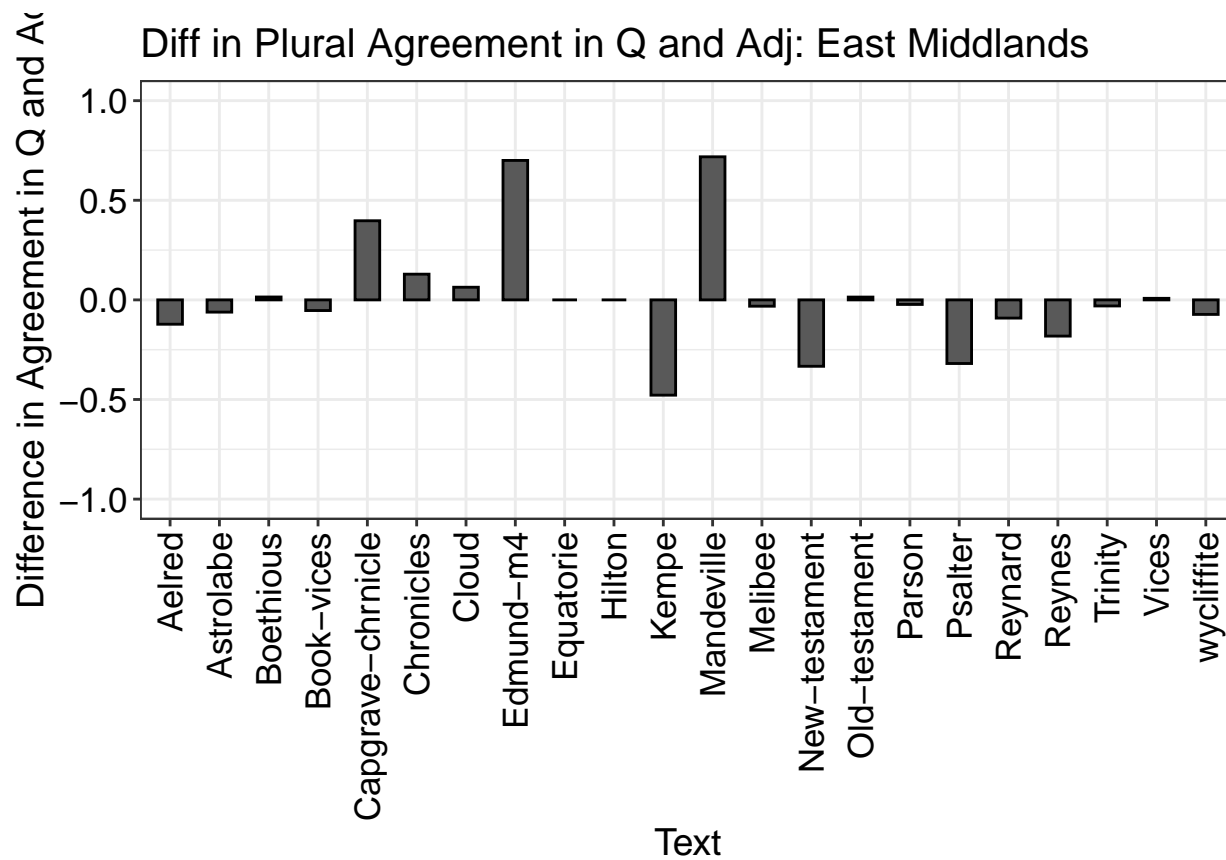
*# Barplots by dialect and Text:*

```
barplot1 <- ggplot(data = total5, aes(x = Text, y = diff)) + geom_bar(stat = "identity") +  
  facet_grid(. ~ Dialect, scales = "free", space = "free") + theme(axis.text.x = element_text(angle = 90,  
    vjust = 0.5, hjust = 1))
```

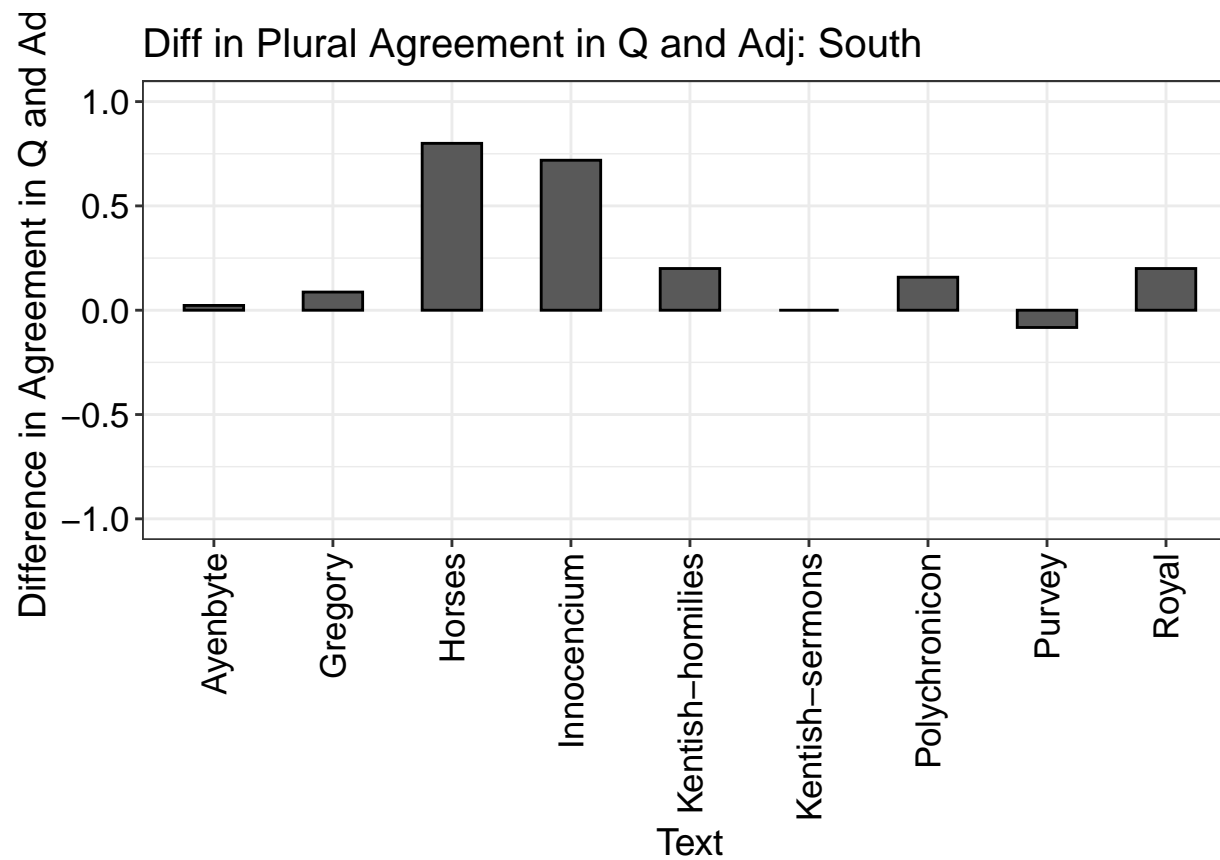
```
barplot1
```



```
# Individual plots
total5EM <- subset(total5, Dialect == "EM")
barplotEM <- ggplot(data = total5EM, aes(x = Text, y = diff)) + geom_bar(stat = "identity",
  width = 0.5) + theme(axis.text.x = element_text(angle = 90, vjust = 0.5, hjust = 1)) +
  ggtitle("Diff in Plural Agreement in Q and Adj: East Midlands") + ylab("Difference in Agreement in Q and Adj") +
  ylim(-1, 1)
barplotEM
```

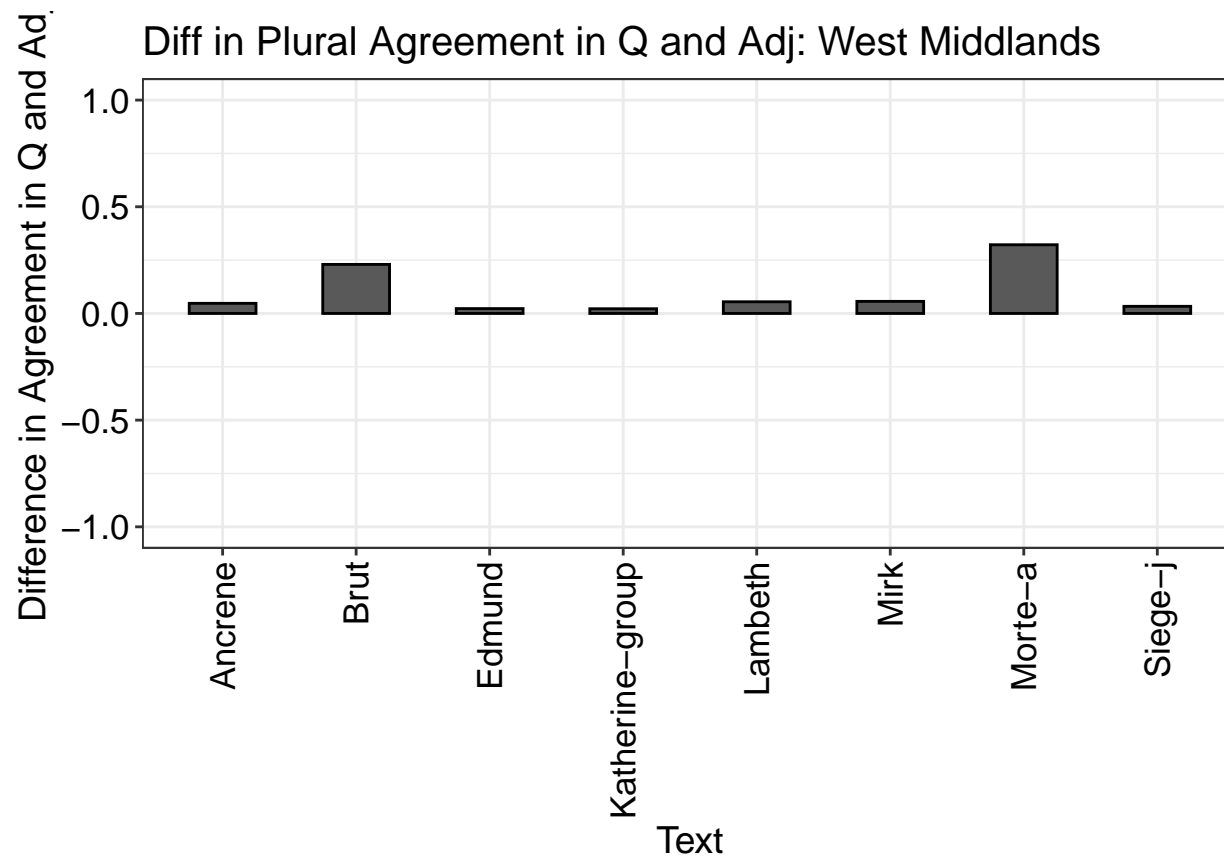


```
#####
total5S <- subset(total5, Dialect == "Southern")
barplotS <- ggplot(data = total5S, aes(x = Text, y = diff)) + geom_bar(stat = "identity",
  width = 0.5) + theme(axis.text.x = element_text(angle = 90, vjust = 0.5, hjust = 1)) +
  ggtitle("Diff in Plural Agreement in Q and Adj: South") + ylab("Difference in Agreement in Q and Adj") +
  ylim(-1, 1)
barplotS
```

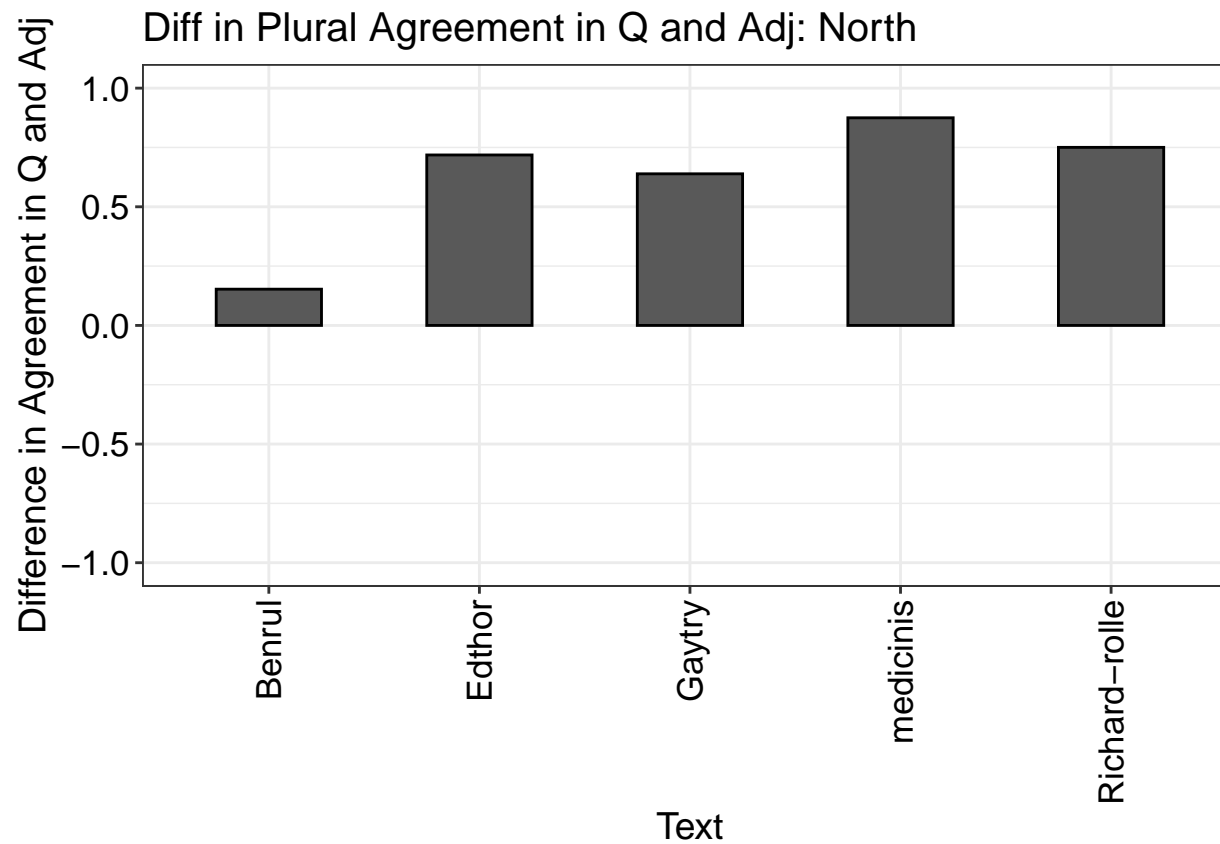


```
#####
total5WM <- subset(total5, Dialect == "WM")
barplotWM <- ggplot(data = total5WM, aes(x = Text, y = diff)) + geom_bar(stat = "identity",
  width = 0.5) + theme(axis.text.x = element_text(angle = 90, vjust = 0.5, hjust = 1)) +
  ggtitle("Diff in Plural Agreement in Q and Adj: West Midlands") + ylab("Difference in Agreement in Q and Adj") +
  ylim(-1, 1)
barplotWM
```





```
total5N <- subset(total5, Dialect == "Northern")
barplotN <- ggplot(data = total5N, aes(x = Text, y = diff)) + geom_bar(stat = "identity",
  width = 0.5) + theme(axis.text.x = element_text(angle = 90, vjust = 0.5, hjust = 1)) +
  ggtitle("Diff in Plural Agreement in Q and Adj: North") + ylab("Difference in Agreement in Q and Adj") +
  ylim(-1, 1)
barplotN
```



```
totalm <- total5 %>%
  mutate(Period = case_when((Text == "medicinis") ~ "M4", (Text == "Horses") ~
    "M3", (Text == "Richard-rolle") ~ "M2", (Text == "Norwich") ~ "M3", (Text ==
    "Innocencium") ~ "M4", (Text == "Edthor") ~ "M3", (Text == "Mandeville") ~
    "M3", (Text == "Edmund-m4") ~ "M4", (Text == "Gaytry") ~ "M3", (Text == "Capgrave-chnicle") ~
    "M4", (Text == "Capgrave-sermon") ~ "M4", (Text == "Morte-a") ~ "M4", (Text ==
    "Brut") ~ "M3", (Text == "Kentish-homilies") ~ "M1", (Text == "Royal") ~
    "M3", (Text == "Polychronicon") ~ "M3", (Text == "Benrui") ~ "M3", (Text ==
    "Chronicles") ~ "M1", (Text == "Gregory") ~ "M4", (Text == "Cloud") ~ "M3",
    (Text == "Mirk") ~ "M3", (Text == "Lambeth") ~ "M1", (Text == "Ancrone") ~
    "M1", (Text == "Siege-j") ~ "M4", (Text == "Ayenbyte") ~ "M2", (Text ==
    "Edmund") ~ "M3", (Text == "Katherine-group") ~ "M1", (Text == "Boethious") ~
```

```

"M3", (Text == "Old-testament") ~ "M3", (Text == "Vices") ~ "M1", (Text ==
"Equatorie") ~ "M3", (Text == "Hilton") ~ "M3", (Text == "Kentish-sermons") ~
"M2", (Text == "Parson") ~ "M3", (Text == "Trinity") ~ "M1", (Text ==
"Melibee") ~ "M3", (Text == "Book-vices") ~ "M3", (Text == "Astrolabe") ~
"M3", (Text == "wycliffite") ~ "M3", (Text == "Purvey") ~ "M3", (Text ==
"Reynard") ~ "M4", (Text == "Aelred") ~ "M2", (Text == "Reynes") ~ "M4",
(Text == "Psalter") ~ "M2", (Text == "New-testament") ~ "M3", (Text == "Kempe") ~
"M4", ))

```

## 4 Statistical Analysis

### 4.1 Linear Regression Model (lm)

```

total$diff <- as.numeric(total$diff)

# Testing the effect of difference in agreement: linear model
ex.fittotal <- lm(diff ~ Dialect, data = total)
summary(ex.fittotal)

##
## Call:
## lm(formula = diff ~ Dialect, data = total)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.48938 -0.08868 -0.04176  0.06209  0.70709
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.01113    0.05720   0.195   0.8467
## DialectNorthern  0.61595    0.13291   4.634 3.77e-05 ***
## DialectSouthern  0.22270    0.10615   2.098  0.0423 *
## DialectWM        0.08744    0.11076   0.789  0.4345
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.2683 on 40 degrees of freedom
## Multiple R-squared:  0.3635, Adjusted R-squared:  0.3158

```

```
## F-statistic: 7.616 on 3 and 40 DF, p-value: 0.0003837
# using South as the reference level:
total$Dialect <- relevel(factor(total$Dialect), ref = 3)
ex.fittotal <- lm(diff ~ Dialect, data = total)
summary(ex.fittotal)

##
## Call:
## lm(formula = diff ~ Dialect, data = total)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.48938 -0.08868 -0.04176  0.06209  0.70709
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.23382    0.08942   2.615  0.0125 *
## DialectEM      -0.22270    0.10615  -2.098  0.0423 *
## DialectNorthern 0.39326    0.14964   2.628  0.0121 *
## DialectWM      -0.13526    0.13036  -1.038  0.3057
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.2683 on 40 degrees of freedom
## Multiple R-squared:  0.3635, Adjusted R-squared:  0.3158
## F-statistic: 7.616 on 3 and 40 DF, p-value: 0.0003837
```

## 4.2 Test of residuals (DHARMA)

```
# DHARMA package:

simulationOutput <- simulateResiduals(fittedModel = ex.fittotal, plot = F)
residuals(simulationOutput)

## [1] 0.268 0.436 0.408 0.220 0.052 0.492 0.380 0.704 0.928 0.644 0.592 0.348
## [13] 0.992 0.660 0.504 0.536 0.256 0.500 0.964 0.968 0.360 0.040 0.472 0.208
## [25] 0.420 0.996 0.836 0.404 0.412 0.848 0.068 0.524 0.444 0.380 0.136 0.120
## [37] 0.308 0.216 0.680 0.456 0.436 0.392 0.456 0.348
```

```
plot(simulationOutput)
```

## DHARMA residual

