The Development of Polarity Subjunctive

Raquel Montero Estebaranz

29.05.2024

Contents

Introduction

This file shows the code used to analyze the data of the verb 'believe' that appears in Chapter 5 of the thesis.

Packages

The following are the packages that will be used throughout this document:

```
library(readr)
library(carData)                              # for cat package
library(car)                                  # Anova function
library(plyr)
library(dplyr, warn.conflicts = FALSE)        # Operations
library(ggplot2)                              # to use ggplot
suppressPackageStartupMessages(library(sjPlot))   # to change the font
library(Matrix)                               # for lme4 package
library(lme4)                                 # to calculate lmer models
library(lmerTest, warn.conflicts = FALSE)     # for the p values
library(ggpubr, warn.conflicts = FALSE)       #multiple plots into same graphs
```

The Development of Mood

Data cleaning

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

```
# Importing the data:
datamood <- read_csv("DiachronyAllVerbs.csv", show_col_types = FALSE,,locale = locale(encoding =   "ISO-8

#Eliminating variables not used:

ex.datamood.believe <- subset(datamood,
                    Mverbtype =="ind" &
                     Mverbclass!="Na"& # not one of the four verbs
                     Emood !="Na" &
                     Emood !="NA" &
                     Emood !="inf" &
                     MClauseType=="noninterrogative" &
                     EClauseType=="unambiguous"&
                     Mverb1=="creer (believe)"
                 )


# Converting Mood in the embedded clause into a 1 and 0:
ex.datamood.believe$Emood2 <- ifelse(ex.datamood.believe$Emood == "subj", 0, 1)
# Converting into a numeric value:
ex.datamood.believe$Emood2 <- as.numeric(as.character(ex.datamood.believe$Emood2))

#relabeling construction:
ex.datamood.believe$Mverbtype2 <- ifelse(ex.datamood.believe$Mverbclass == "factive", "Semi-factive", "
ex.datamood.believe$Construction <- ifelse(ex.datamood.believe$Construction == "1st present", "1st pres
```

Data visualization

```
# Calculating mean values for plotting:
plot.datamood <- ddply(ex.datamood.believe, .(Period, Construction
),summarize, whet = mean(as.numeric(as.character(Emood2)), na.rm = T), n = sum(!is.na(as.numeric(as.char



p <- ggplot(plot.datamood, aes(Period, whet, color=Construction))+
  geom_point(aes(size = n),alpha = 0.9)+
  scale_size_area(max_size=20,limits=c(1,1100))+
  labs(title="Proportion mood: not believe(ind) that p(ind/subj)",
       y="Proportion embedded mood (1=ind, 0=subj)",
       x="",
       size="size",
       color="Construction") +
  scale_color_manual(values=c("black", "#FE6100"))+ # selecting colors manually
  ylim(0,1)#+


p
```

Proportion mood: not believe(ind) that p(ind/subj)
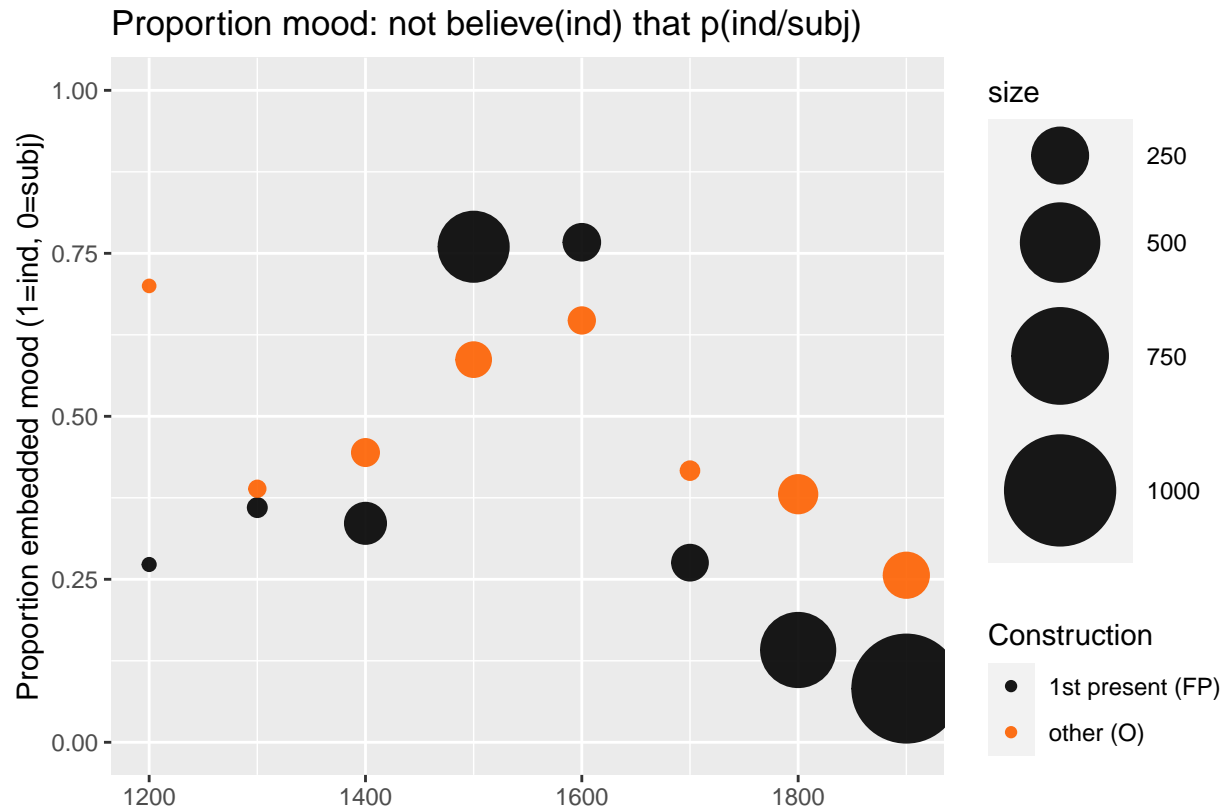

```
ggsave(p, file="plot-believe.png", width = 6.5, height = 4)
```

Next, we plot the same graph but now with geom_smooth and the confidence bands to have a first look as to whether the two constructions might be statistically different:

```
psmooth <- ggplot(ex.datamood.believe, aes(Period,Emood2 , color=Construction))+
  geom_point(data=plot.datamood, aes(Period, whet, color=Construction, size=n), alpha=0.8)+
  scale_size_area(max_size=20,limits=c(1,1100))+
  labs(title="Proportion mood: not believe(ind) that p(ind/subj)",
       y="Proportion embedded mood (1=ind, 0=subj)",
       size="Amount of data",
       color="Construction") +
  geom_smooth(method="loess",se=TRUE,linewidth=0.5, fill="gray85")+
  coord_cartesian(ylim = c(0,1))+  # lets the confidence bands on even if they go below 0
  scale_color_manual(values=c("black", "#FE6100")) # selecting colors manually
psmooth
```

Proportion mood: not believe(ind) that p(ind/subj)
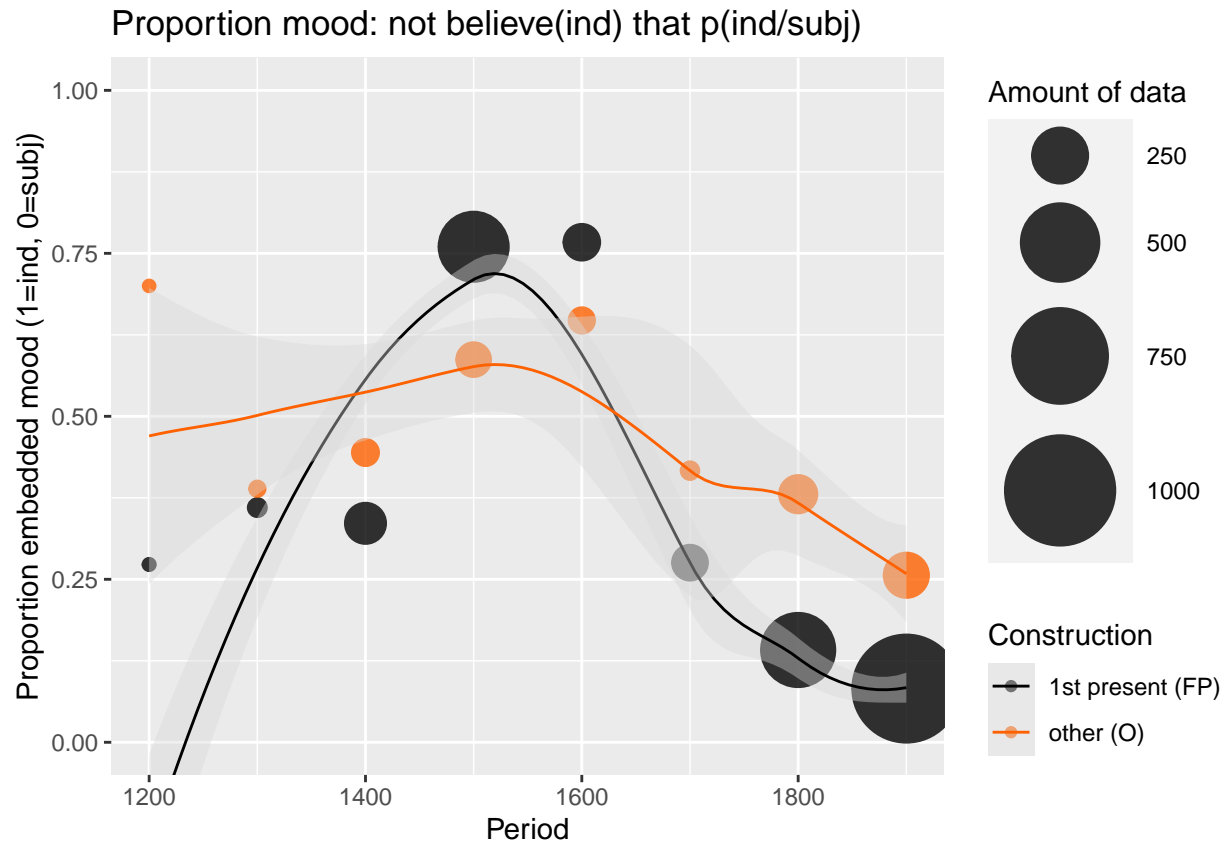
In the key periods the confidence bands do not overlap: 1500 and 1800-1900, and hence the data seems to suggest that indeed there is a difference between the two constructions, and that there was a shift in the proportion of mood each constructions took: 1500 1st person took more indicative, but later in time 1800-1900 1st person took less subjunctive.

```
#Saving the plots:
ggsave(psmooth,file="moodsmooth2.png", width=8, height = 5)
```

Fitting the Laplace function

Given that the structures are different, we want then to calculate their speeds of change. In order to do so, we can fit the Laplace curve to our data set:

```
#Define a function to fit data to:
laplace <- function(Period,m,k,s)(m*exp(-(abs(Period-k)/s)))
```

First person

```
#subsetting the data:
ex.datamoodfirst <- subset(ex.datamood.believe, Construction=="1st present (FP)")
#NLLS:
model <- nls(Emood2 ~ laplace(Period,m,k,s), data=ex.datamoodfirst, start=list(m=0.43,k=1572, s=200))
summary(model) #m=1.1, k=1548, s=128
```

```
##
## Formula: Emood2 ~ laplace(Period, m, k, s)
##
```

```
## Parameters:
##   Estimate Std. Error t value Pr(>|t|)
## m 1.108e+00  4.268e-02    25.97    <2e-16 ***
## k 1.548e+03  3.181e+00   486.67    <2e-16 ***
## s 1.279e+02  6.224e+00    20.56    <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.3571 on 2165 degrees of freedom
##
## Number of iterations to convergence: 8
## Achieved convergence tolerance: 3.931e-06
```

The model goes above 1 so we need to add a constraint so that q does not reach past 1:

```
modelbis <- nls(Emood2 ~ laplace(Period,m,k,s), data=ex.datamoodfirst, start=list(m=0.43,k=1572, s=200)
summary(modelbis) #m = 1, k = 1542, s = 142
```

```
##
## Formula: Emood2 ~ laplace(Period, m, k, s)
##
## Parameters:
##   Estimate Std. Error t value Pr(>|t|)
## m 1.000e+00  4.043e-02    24.74    <2e-16 ***
## k 1.542e+03  3.957e+00   389.67    <2e-16 ***
## s 1.423e+02  7.365e+00    19.32    <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.3576 on 2165 degrees of freedom
##
## Algorithm "port", convergence message: both X-convergence and relative convergence (5)
```

Non-first person

```
ex.datamoodother <- subset(ex.datamood.believe, Construction=="other (O)")
```

```
#The same is done for non-first person data:
modelother <- nls(Emood2 ~ laplace(Period,m,k,s), data=ex.datamoodother, start=list(m=0.7,k=1549,s=370)
summary(modelother)
```

```
##
## Formula: Emood2 ~ laplace(Period, m, k, s)
##
## Parameters:
##   Estimate Std. Error t value Pr(>|t|)
## m 6.835e-01  5.712e-02   11.966  < 2e-16 ***
## k 1.551e+03  2.139e+01   72.521  < 2e-16 ***
## s 3.981e+02  7.695e+01    5.174 3.28e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.4783 on 519 degrees of freedom
##
```

```
## Algorithm "port", convergence message: relative convergence (4)
#m=0.68, k=1551, s=398
```

Plot

In order to check if the obtained curves are a good fit for the data, we plot them next to the data:

```
fitB <- function(x){0.68* exp(-(abs(x-1551)/398))}
fitA <- function(x) {1* exp(-(abs(x-1542)/142))}

plot.datamood$Period <- as.numeric(as.character(plot.datamood$Period))

pLaplace <- ggplot(plot.datamood, aes(Period, whet, color=Construction))+
  geom_point(aes(size = n),alpha = 0.9)+
  scale_size_area(max_size=20,limits=c(1,1100))+
  labs(title="Proportion mood: not believe(ind) that p(ind/subj)",
       y="Proportion embedded mood (1=ind, 0=subj)",
       size="size",
       x="",
       color="Construction") +
  scale_color_manual(values=c("black", "#FE6100")) +
  ylim(0,1.01) +
  geom_function(fun=fitA, color = "black")+
 geom_function(fun=fitB, color = "#FE6100")

pLaplace
```


Proportion mood: not believe(ind) that p(ind/subj)
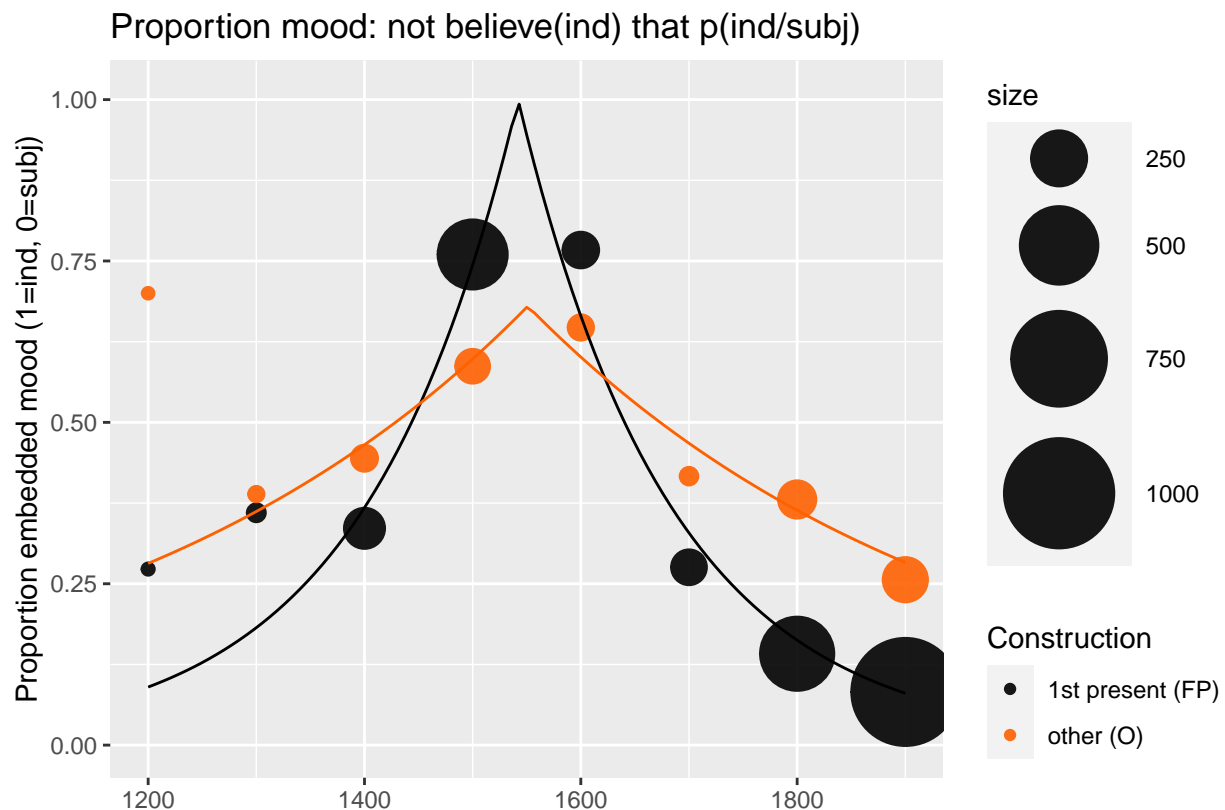
```r
ggsave(pLaplace, file = "MoodR3.png", width = 8, height = 5)
```

Adding the confidence intervals

```r
#Main functions:
fitB <- function(x){0.68* exp(-(abs(x-1551)/398))}
fitA <- function(x) {1* exp(-(abs(x-1542)/142))}

#Confidence intervals for the speed of Construction O:
fitBupper <-function(x){0.8* exp(-(abs(x-1593)/552))}
fitBlower <-function(x){0.56* exp(-(abs(x-1509)/244))}
#Confidence intervals for the speed of construction FP:
fitAupper <- function(x) {1.08* exp(-(abs(x-1550)/156))}
fitAlower <- function(x) {0.92* exp(-(abs(x-1534)/128))}

#equations for plotting the confiddence intervals:
x <- seq(1200,1900,10)
y1O <- fitBupper(x)
y2O <- fitBlower(x)

x <- seq(1200,1900,10)
y1FP <- fitAupper(x)
y2FP <- fitAlower(x)



plot.datamood$Period <- as.numeric(as.character(plot.datamood$Period))

pLaplace.ci <- ggplot()+
  geom_point(data=plot.datamood,
                        aes(Period, whet, size = n, color=Construction))+
  scale_size_area(max_size=20,limits=c(1,1100))+
  labs(title="Proportion mood: not believe(ind) that p(ind/subj)",
       y="Proportion embedded mood (1=ind, 0=subj)",
       size="size",
       x="",
       color="Construction") +
  scale_color_manual(values=c("black", "#FE6100")) +
  ylim(0,1.10) +
  geom_function(fun=fitA, color = "black")+
 geom_function(fun=fitB, color = "#FE6100")+
  geom_polygon(aes(c(x,rev(x)),c(y2O,rev(y1O))),fill="#FE6100",alpha=0.1)+
  geom_polygon(aes(c(x,rev(x)),c(y2FP,rev(y1FP))),fill="gray",alpha=0.3)

pLaplace.ci
```
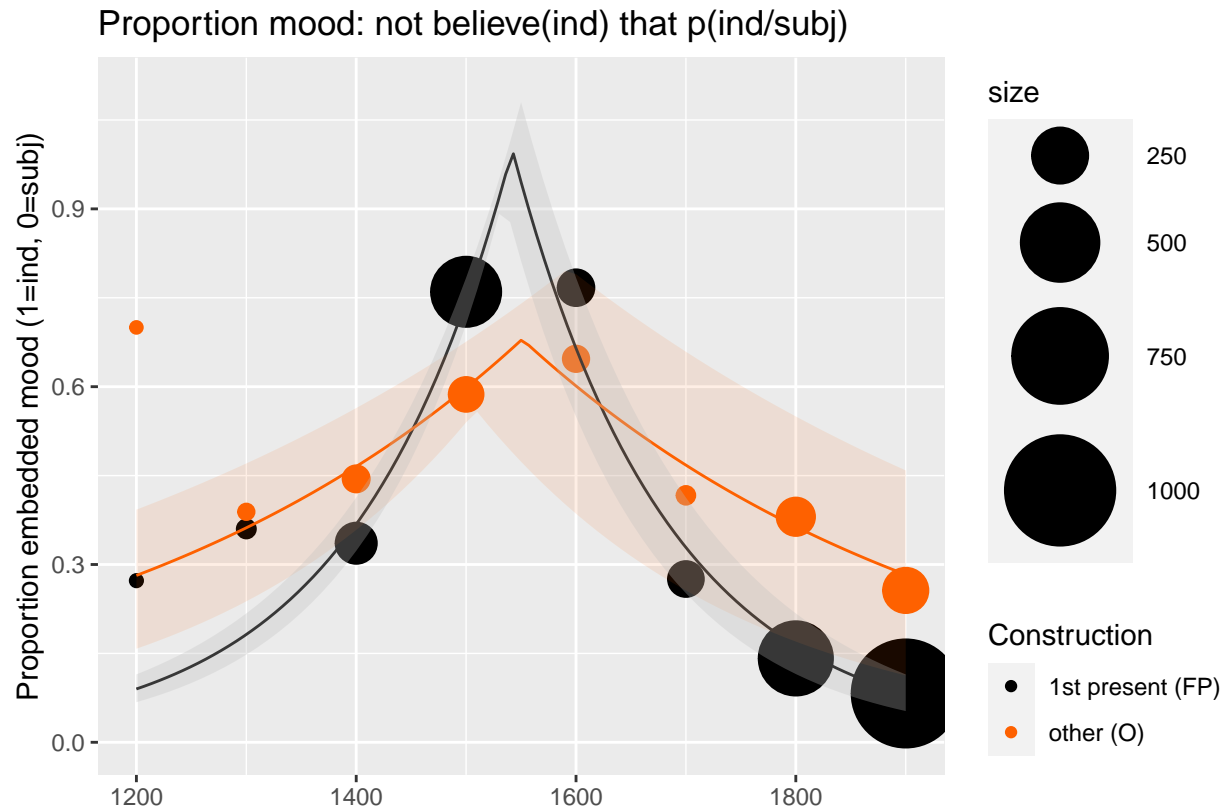
Proportion mood: not believe(ind) that p(ind/subj)

```
ggsave(pLaplace.ci, file = "believe-mood-fit-ci.png", width = 8, height = 5)
```

The Development of Complementizers

```
datac <- read_csv("ComplementizersAnnotated.csv", show_col_types = FALSE,locale = locale(encoding =  "IS
```

Data cleaning

```
ex.datac <- subset(datac,
                   ClauseType =="non-interrogative" &
                   Structure != "other"&
                   MatrixMood=="ind" &
                   Everbfiniteness=="finite"
                  )

ex.datacsubj <- subset(datac,
                   ClauseType =="non-interrogative" &
                   Structure != "other"&
                   MatrixMood=="ind" &
                   Everbfiniteness=="finite" &
                   Everbmood=="subj"
                  )

ex.datacind <- subset(datac,
```

```r
                        ClauseType =="non-interrogative" &
                        Structure != "other"&
                        MatrixMood=="ind" &
                        Everbfiniteness=="finite" &
                        Everbmood=="ind"
                         )


ex.datacfirstind <- subset(ex.datacind,
                            Construction=="1st present (creo, creemos)"
                          )


ex.datacotherind <- subset(ex.datacind,
                             Construction!="1st present (creo, creemos)"
                           )

ex.datacsubj$Complementizer2 <- ifelse(ex.datacsubj$Complementizer == "That", 0, 1)
ex.datacsubj$Complementizer2 <- as.numeric(as.character(ex.datacsubj$Complementizer2))

ex.datacind$Complementizer2 <- ifelse(ex.datacind$Complementizer == "That", 0, 1)
ex.datacind$Complementizer2 <- as.numeric(as.character(ex.datacind$Complementizer2))
ex.datacind$Construction <- ifelse(ex.datacind$Construction == "1st present (creo, creemos)", "1st prese

ex.datac$Complementizer2 <- ifelse(ex.datac$Complementizer == "That", 0, 1)
ex.datacComplementizer2 <- as.numeric(as.character(ex.datac$Complementizer2))
ex.datac$Construction <- ifelse(ex.datac$Construction == "1st present (creo, creemos)", "1st present (Fl


ex.datacfirstind$Complementizer2 <- ifelse(ex.datacfirstind$Complementizer == "That", 0, 1)
ex.datacfirstind$Complementizer2 <- as.numeric(as.character(ex.datacfirstind$Complementizer2))

ex.datacotherind$Complementizer2 <- ifelse(ex.datacotherind$Complementizer == "That", 0, 1)
ex.datacotherind$Complementizer2 <- as.numeric(as.character(ex.datacotherind$Complementizer2))

# Data for Plotting

plot.datac <- ddply(ex.datac, .(Period, Construction, Everbmood
),summarize, whet = mean(as.numeric(as.character(Complementizer2)), na.rm = T), n = sum(!is.na(as.numeri

plot.datacsubj <- ddply(ex.datacsubj, .(Period, Construction
),summarize, whet = mean(as.numeric(as.character(Complementizer2)), na.rm = T), n = sum(!is.na(as.numeri

plot.datacind <- ddply(ex.datacind, .(Period, Construction
),summarize, whet = mean(as.numeric(as.character(Complementizer2)), na.rm = T), n = sum(!is.na(as.numeri


plot.datacfirstind <- ddply(ex.datacfirstind, .(Period),summarize, whet = mean(as.numeric(as.character(
```
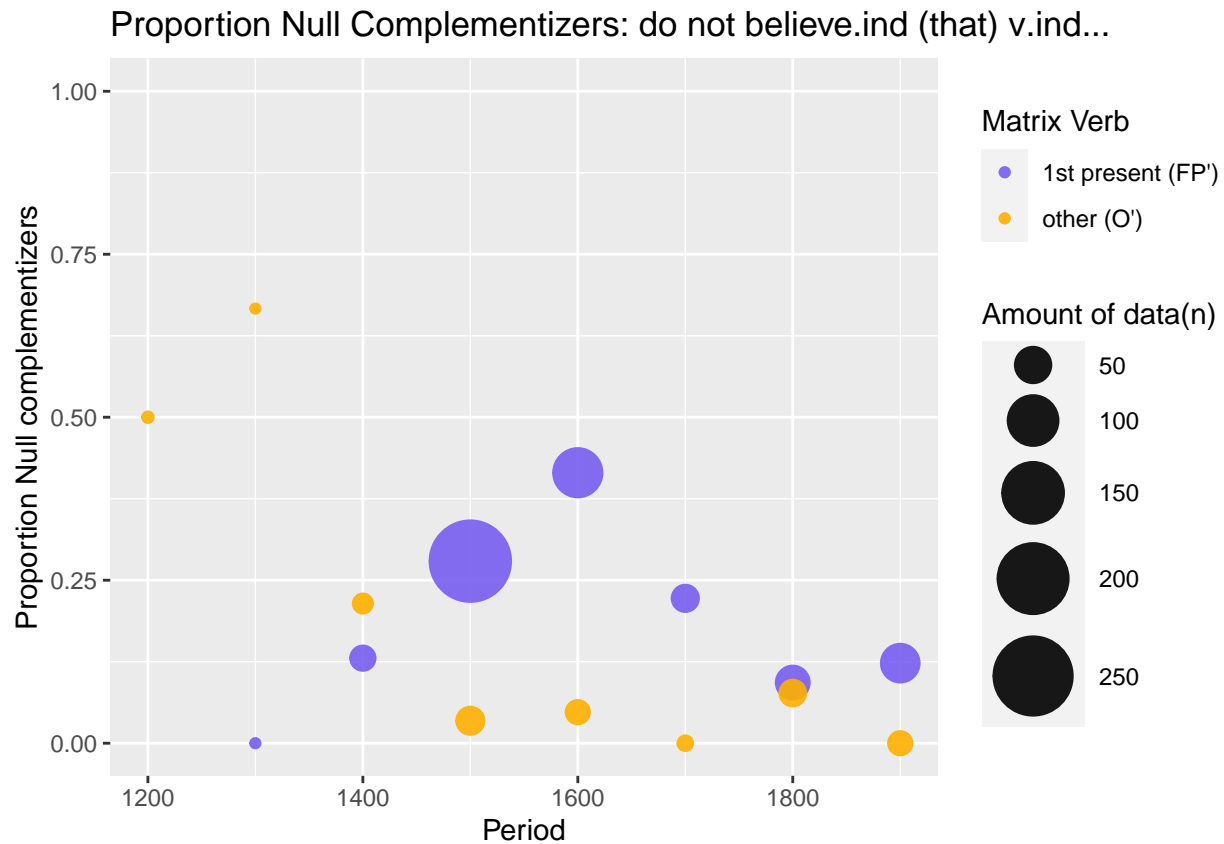
Data visualization

```r
pind <- ggplot(plot.datacind, aes(Period, whet, color=Construction))+
  geom_point(aes(size = n),alpha = 0.9)+
```

```
  scale_size_area(max_size=14)+
  labs(title="Proportion Null Complementizers: do not believe.ind (that) v.ind...", y="Proportion Null
  scale_fill_discrete(labels=c('No Vind that', 'No Vsubj that')) +
  scale_color_manual(values=c("#785EF0", "#FFB000"))  +
  ylim(0,1)
pind
```

Proportion Null Complementizers: do not believe.ind (that) v.ind...


```
pboth <- ggplot(plot.datac, aes(Period, whet, color=Construction))+
  geom_point(aes(size = n),alpha = 0.9)+
  scale_size_area(max_size=14)+
  facet_wrap(~Everbmood,ncol=2)+
  labs(title="Proportion Null Complementizers: do not believe.ind (that) v.ind/subj...", y="Proportion N
  scale_fill_discrete(labels=c('No Vind that', 'No Vsubj that')) +
  scale_color_manual(values=c("#785EF0", "#FFB000"))  +
  ylim(0,1)
#pboth
```

We then plot the data with a smooth calculated using the loess method.
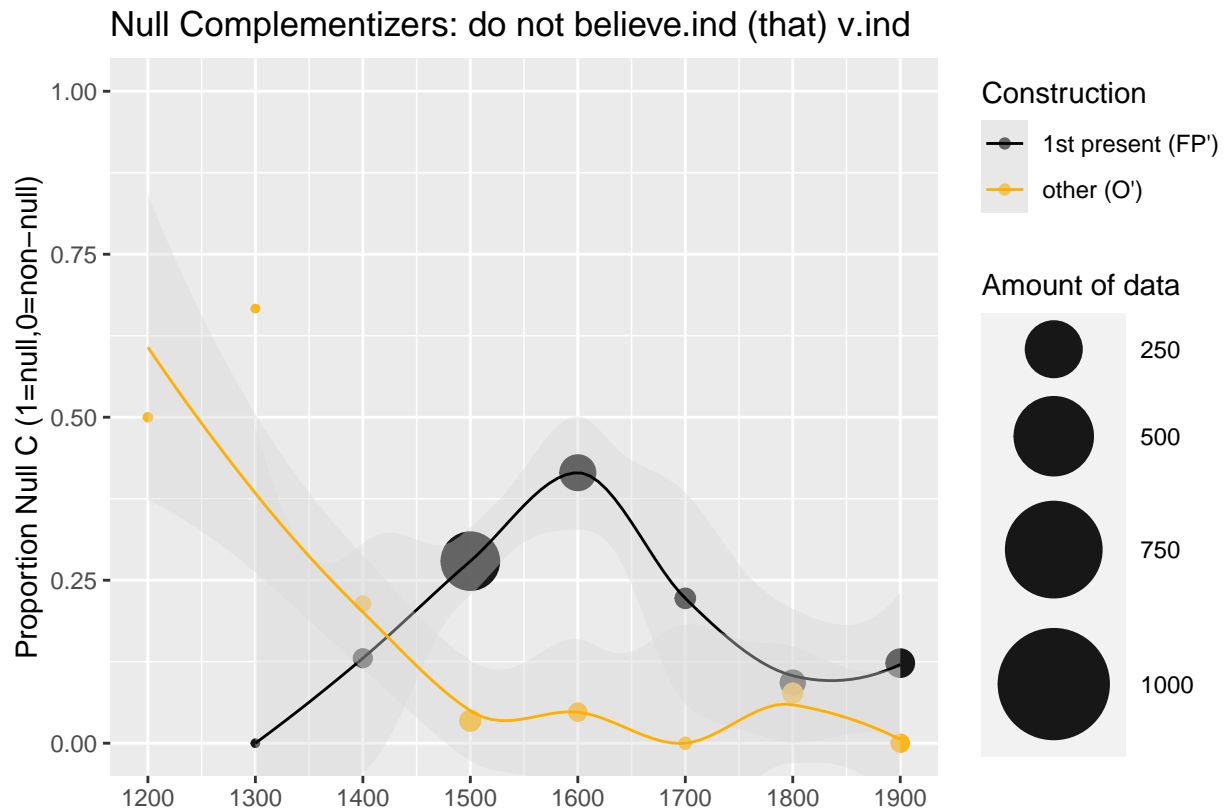
```
pindsmooth <- ggplot(ex.datacind, aes(Period, Complementizer2, color=Construction))+
  geom_point(data=plot.datacind, aes(Period, whet, color=Construction, size = n),alpha = 0.9)+
  scale_size_area(max_size=20,limits=c(1,1100))+
  labs(title="Null Complementizers: do not believe.ind (that) v.ind",
       y="Proportion Null C (1=null,0=non-null)",
       x="",
       size="Amount of data",
       color="Construction")+
```

```
    scale_fill_discrete(labels=c('No Vind that', 'No Vsubj that'))  +
    geom_smooth(method = loess, se=TRUE, fill="gray85", linewidth=0.5)+
    coord_cartesian(ylim=c(0,1))+
    scale_x_continuous(breaks=seq(1200,1900,100))+
    scale_color_manual(values=c("black", "#FFB000")) # selecting colors manually
pindsmooth
```


Null Complementizers: do not believe.ind (that) v.ind

```
pbothsmooth <- ggplot(ex.datac, aes(Period, Complementizer2, whet, color=Construction))+
    geom_point(data=plot.datac, aes(Period, whet, color=Construction, size = n),alpha = 0.9)+
    scale_size_area(max_size=20,limits=c(1,1100))+
    facet_wrap(~Everbmood,ncol=2)+
    labs(title="Proportion Null Complementizers: do not believe.ind (that) v.ind/subj...",
        y="Proportion Null complementizers",
        size="Amount of data",
        color="Construction",
        x="")+
    scale_fill_discrete(labels=c('No Vind that', 'No Vsubj that')) +
    geom_smooth(method = loess, se=TRUE, fill="gray85", linewidth=0.5)+
    coord_cartesian(ylim=c(0,1))+
    scale_x_continuous(breaks=seq(1200,1900,100))+
    scale_color_manual(values=c("black", "#FFB000"))
#pbothsmooth

ggsave(pbothsmooth, file="complementizersSmooth.png", width = 8, height = 4)
```

Idea: multiple effects

```r
psmooth <- ggplot(ex.datamood.believe, aes(Period,Emood2 , color=Construction))+
  geom_point(data=plot.datamood, aes(Period, whet, color=Construction, size=n), alpha=0.8)+
  scale_size_area(max_size=20,limits=c(1,1100))+
  labs(title="Proportion mood: not believe(ind) that p(ind/subj)",
       y="Proportion embedded mood (1=ind, 0=subj)",
       size="Amount of data",
       color="Construction") +
  geom_smooth(method="loess",se=TRUE,linewidth=0.5, fill="gray85")+
  coord_cartesian(ylim = c(0,1))+  # lets the confidence bands on even if they go below 0
  scale_color_manual(values=c("black", "#FE6100"))+ # selecting colors manually
   annotate("text", x = 1600, y = 0.9,
                       label = "Change MFP = Change M + Change C",
                       family= theme_get()$text[["family"]],
                        size= theme_get()$text[["size"]]/2.5,
                       color="black"
                     )+
 annotate("text", x = 1800, y = 0.7,
                       label = "Change MO = Change M",
                       family= theme_get()$text[["family"]],
                        size= theme_get()$text[["size"]]/2.5,
                       color="#FE6100"
                     )
psmooth
```


Proportion mood: not believe(ind) that p(ind/subj)
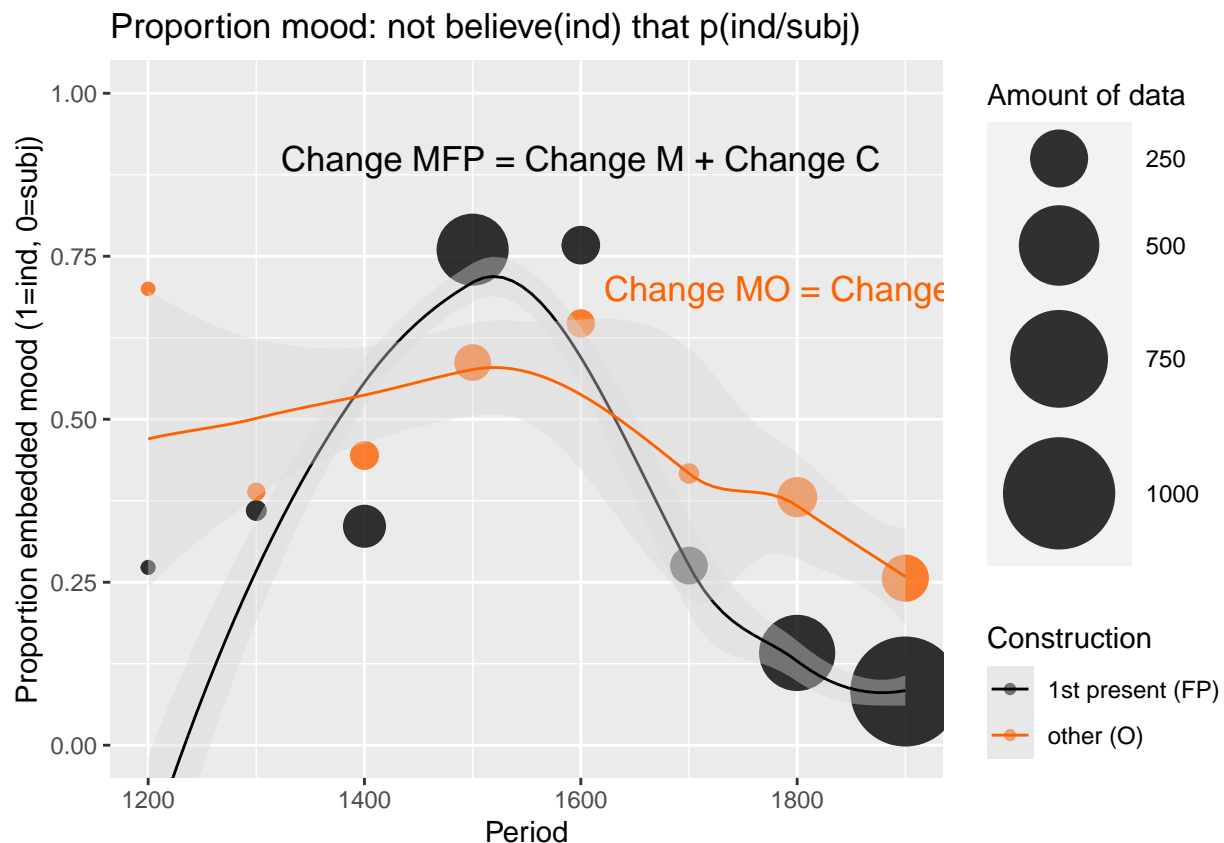
```
pcfirst <- ggplot(ex.datacfirstind, aes(Period,Complementizer2))+
  geom_point(data=plot.datacfirstind, aes(Period, whet, size=n), alpha=0.8)+
  scale_size_area(max_size=20,limits=c(1,1100))+
  labs(title="Proportion mood: I not believe(ind) that p(ind/subj)",
       y="Proportion embedded mood (1=ind, 0=subj)",
       size="Amount of data",
       color="Construction") +
  geom_smooth(method="loess",se=TRUE,linewidth=0.5, fill="gray85")+
  coord_cartesian(ylim = c(0,1))+  # lets the confidence bands on even if they go below 0
  annotate("text", x = 1600, y = 0.6,
                      label = "Change CFP = Change C",
                      family= theme_get()$text[["family"]],
                       size= theme_get()$text[["size"]]/2.5,
                      color="blue"
                     )
  #scale_color_manual(values="black") # selecting colors manually
pcfirst
```

Proportion mood: I not believe(ind) that p(ind/subj)
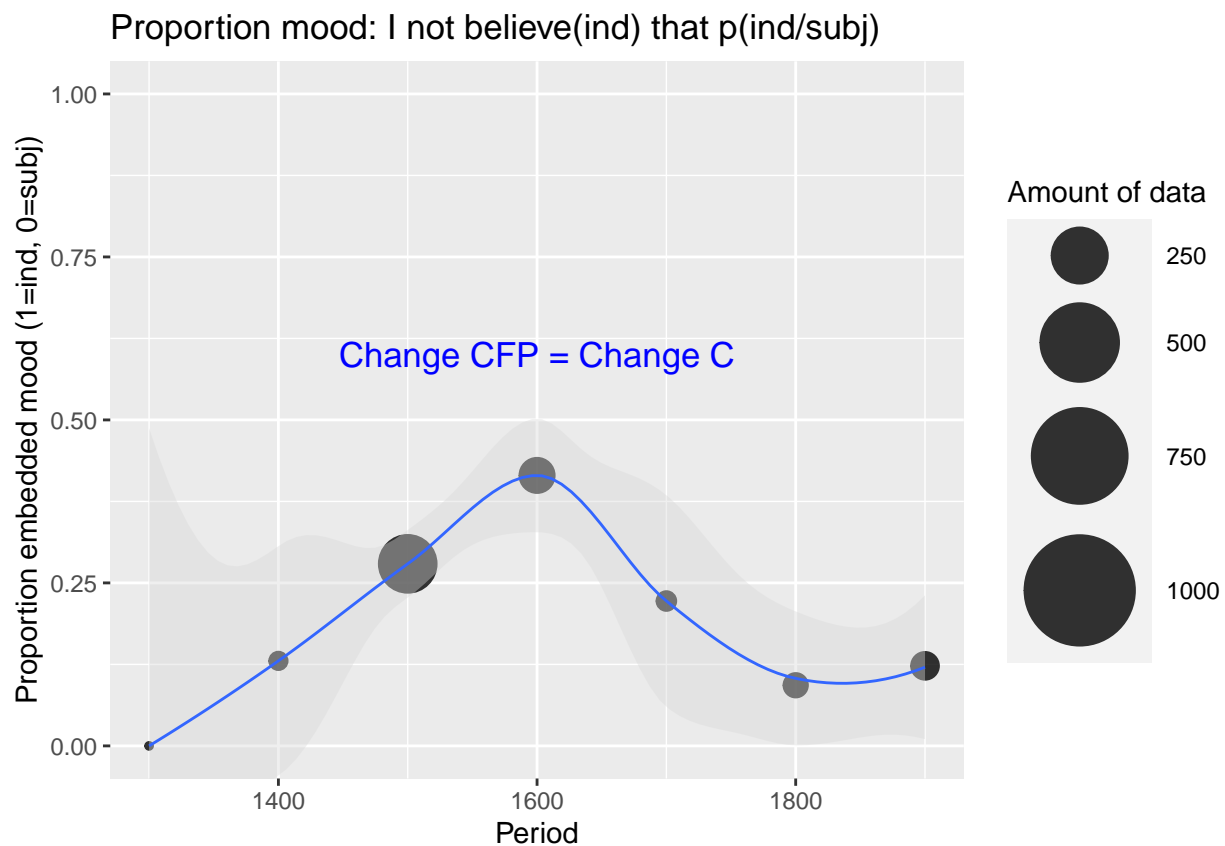

```
AdditionEffectsIdea <- ggarrange(psmooth,pcfirst)
#AdditionEffectsIdea
```

```
ggsave(AdditionEffectsIdea, file="AdditionEffectsIdea.png", width = 15, height = 5)
```

Fitting the Laplace

First person

```
modelc <- nls(Complementizer2 ~ laplace(Period,m,k,s), data=ex.datacfirstind, start=list(m=0.44,k=1589,
summary(modelc) #m=0.45 k=1584 s=174
```

```
##
## Formula: Complementizer2 ~ laplace(Period, m, k, s)
##
## Parameters:
##   Estimate Std. Error t value Pr(>|t|)
## m 4.519e-01  5.479e-02    8.248 1.39e-15 ***
## k 1.584e+03  1.385e+01  114.362  < 2e-16 ***
## s 1.735e+02  4.881e+01    3.554 0.000414 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.4285 on 509 degrees of freedom
##
## Algorithm "port", convergence message: relative convergence (4)
```

Non-first person

For non-first person constructions the fitting gives a singular gradient, indicating that it is not a good fit for the model, and the change cannot be modelled using the Laplace.

```
#The code has been commented for rendering purposes, as it gives a non-singular gradient and wouldn't a
#modelco <- nls(Complementizer2 ~ laplace(Period,m,k,s), data=ex.datacotherind, start=list(m=0.3,k=1700
#summary(modelco)
```

Plot

In order to see if in the case of first person, the obtained curve is a good fit, we plot the data and the curve:

```
pcind3 <- ggplot(plot.datacfirstind, aes(Period, whet))+
  geom_point(aes(size = n),alpha = 0.9, color="black")+
    scale_size_area(max_size=20,limits=c(1,1100))+
  labs(title="Proportion Complementizers: I/we don't believe.ind (that) V.ind ...",
       y="Proportion complementizers (1=null, 0=that)",
       size="Amount of data",
       x="",
       color="Matrix Verb")+
  scale_fill_discrete(labels=c('No Vind that', 'No Vsubj that')) +
  ylim(0,1) +
 geom_function(fun=function(x)0.44* exp(-(abs(x-1583)/178)), color = "black")+
  xlim(1200,1900)
pcind3
```

Proportion Complementizers: I/we don't believe.ind (that) V.ind ...
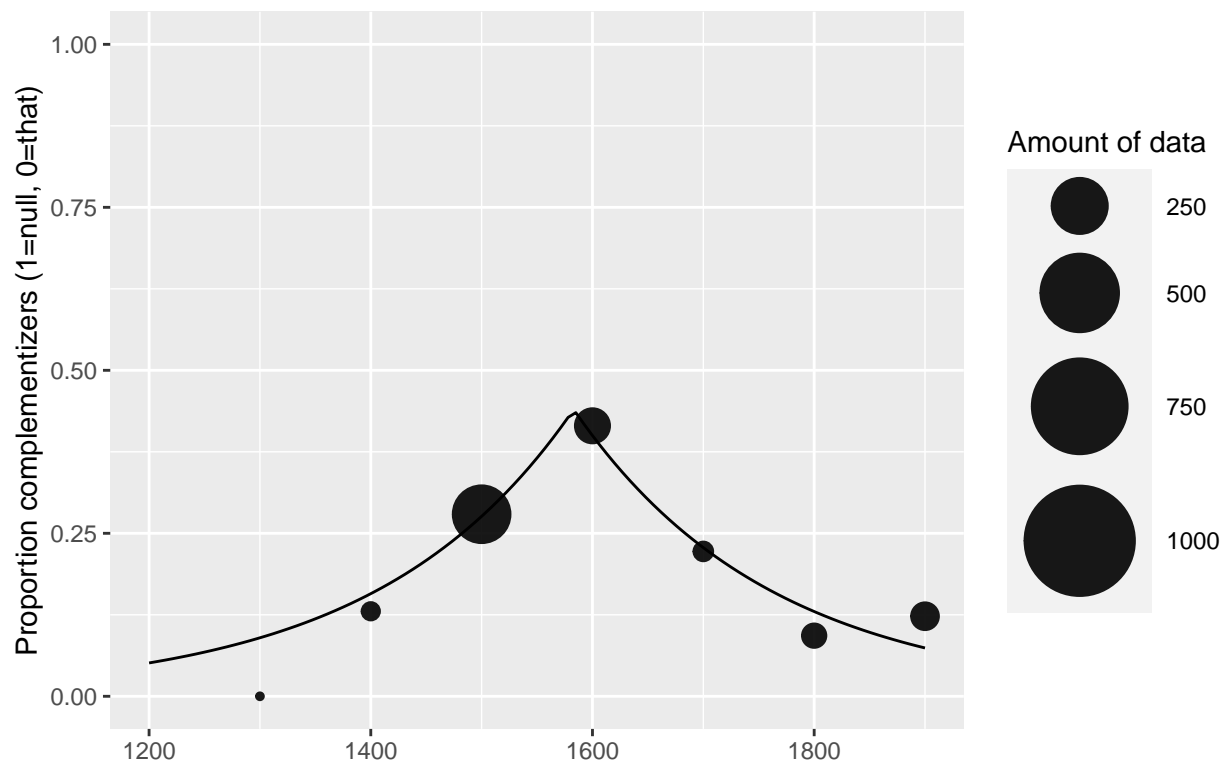


```
fitting <- ggarrange(pLaplace,pcind3)
#fitting
```

```
ggsave(fitting, file="fittingChanges.png", width = 15, height = 5)
ggsave(pcind3, file="complementizerind1st.png", width = 8, height = 5)
```

Adding the Effects

```
#calculating the proportions for only first person:
plot.datamoodfirst <- ddply(ex.datamoodfirst, .(Period),summarize, whet = mean(as.numeric(as.character(

#calculating the proportions for non-first person constructions:
plot.datamoodother <- ddply(ex.datamoodother, .(Period),summarize, whet = mean(as.numeric(as.character(

paddition <- ggplot(plot.datamoodfirst, aes(Period, whet))+
  geom_point(aes(size = n),alpha = 0.9)+
  scale_size_area(max_size=14)+
  labs(title="Proportion mood: not believe(ind) that p(ind/subj)", y="Proportion embedded mood (1=ind,
  geom_function(fun=function(x)1* exp(-(abs(x-1542)/142)),aes(color = "Orignal Model"))+
 geom_function(fun=function(x)1* exp(-(abs(x-1568)/121)),aes(color = "Derived Model"), linetype = "dash
paddition
```

Proportion mood: not believe(ind) that p(ind/subj)
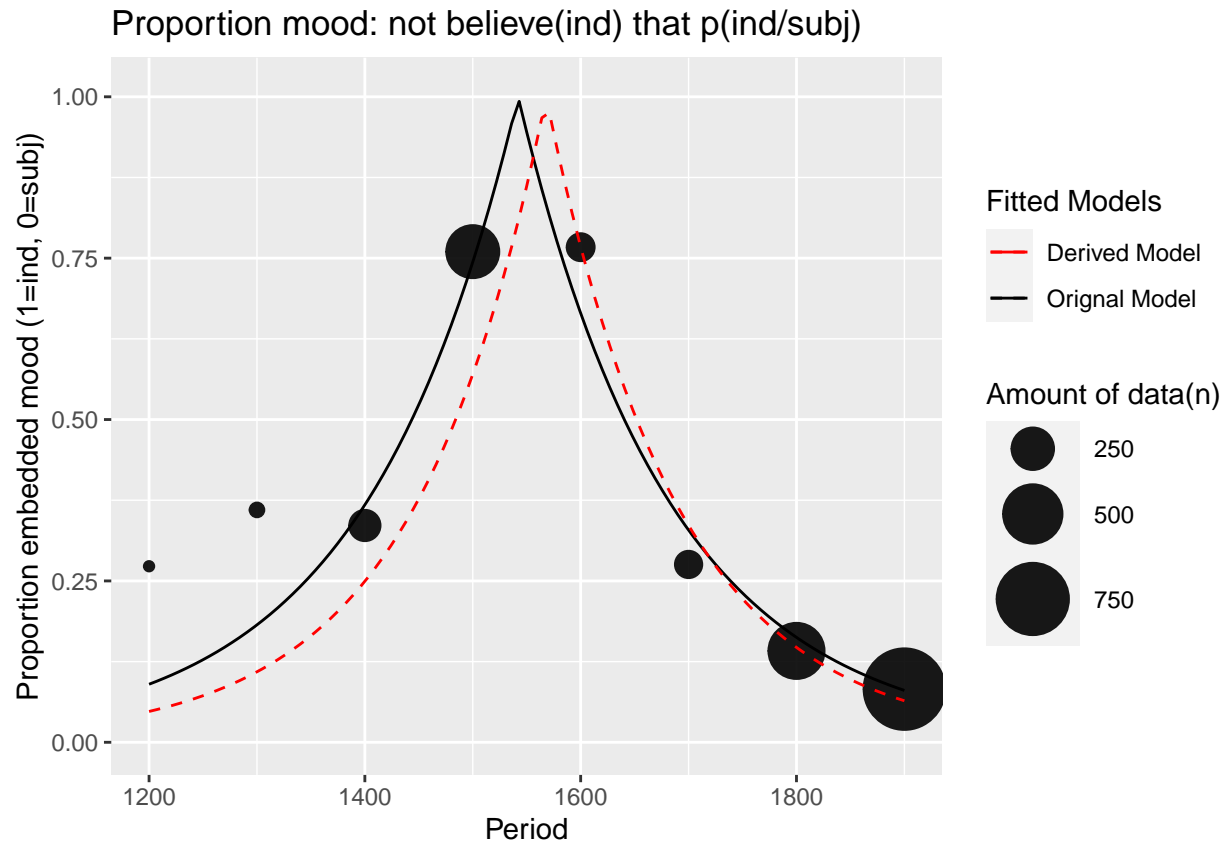
```
ggsave(paddition, file = "addition-effects.png", width = 8, height = 5)
```