Long Before Short Analysis

2024-05-07

Argument Analysis

a. Dialogue Data

1. Load the Data

```
## 'data.frame':    1526 obs. of  28 variables:
##  $ File                   : chr  "../Dialouge Corpus Filttered/parse_gold_filttered\\Phase1\\Phase1_g
##  $ Sent_ID                : int  47 77 79 98 108 129 140 168 181 215 ...
##  $ Sentence               : chr  "              " "                                          " "
##  $ Length                 : int  8 18 14 8 9 7 7 11 11 12 ...
##  $ Word.Order             : chr  "SOV" "SOV" "SOV" "SOV" ...
##  $ Category               : chr  "Transitive" "Transitive" "Transitive" "Transitive" ...
##  $ dependency.length      : num  1.83 2.75 2.17 2.5 2.43 ...
##  $ sub.ID                 : int  1 2 3 5 2 4 3 1 1 6 ...
##  $ subject                : chr  " " " " " " " " " " ...
##  $ sub_subset_words       : chr  "['  ']" "[' ']" "['   ']" "[' ']" ...
##  $ sub_subset_length      : int  1 1 1 1 1 1 1 1 1 1 ...
##  $ sub_subset_char_length : int  5 2 4 3 2 3 3 2 3 4 ...
##  $ sub_difluency          : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ sub_hesitation         : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ sub_repair             : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ obj.ID                 : int  3 5 6 6 6 5 1 5 3 5 ...
##  $ object                 : chr  "  " "  "  "  "  "  " ...
##  $ obj_subset_words       : chr  "['  ']" "['  ']" "['   ', '   ']" "['  ']" ...
##  $ obj_subset.length      : int  1 1 2 1 3 1 2 1 2 4 ...
##  $ obj_subset_char_length : int  5 3 10 4 16 4 8 3 7 21 ...
##  $ obj_difluency          : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ obj_hesitation         : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ obj_repair             : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ iobj.ID                : int  NA NA NA NA NA NA NA NA NA NA ...
##  $ iobject                : chr  "" "" "" "" ...
##  $ iobj_subset_words      : chr  "" "" "" "" ...
##  $ iobj_subset_length     : int  NA NA NA NA NA NA NA NA NA NA ...
##  $ iobj_subset_char_length: int  NA NA NA NA NA NA NA NA NA NA ...
```

2. Processing the Data

```
word_order_data$File_unique<-paste(word_order_data$File, word_order_data$Sent_ID, sep="")

#Three conditions for subj obj length
word_order_data$length_condition<-
```

```
  ifelse(word_order_data$sub_subset_length>word_order_data$obj_subset.length,'S_Long',
  ifelse(word_order_data$sub_subset_length<word_order_data$obj_subset.length,'O_Long','Equal'))

word_order_data$length_condition<-as.factor(word_order_data$length_condition)

#Subset Transitive Data
Transitive_Data <- subset(word_order_data, word_order_data$Category=='Transitive')

#Order: SOV=0 OSV=1
Transitive_Data$Order<- ifelse(Transitive_Data$Word.Order=="SOV",0,1)

#Contrast

# Treatment Contrast
#-------------------
# S-Long   O-Long
# ---------------
#   1        0
#   0        1
#  -1       -1
#-------------------

#Contrast matrix:
#  OLong SLong
#Equal -1/3  -1/3
#OL     2/3  -1/3
#SL    -1/3   2/3

Transitive_Data$S_Long<-ifelse(Transitive_Data$length_condition=="S_Long",2/3,-1/3)
Transitive_Data$O_Long<-ifelse(Transitive_Data$length_condition=="O_Long",2/3,-1/3)


Transitive_Data$dependency.length<- scale(Transitive_Data$dependency.length)

summary(Transitive_Data)
```

```
##      File              Sent_ID         Sentence            Length
##  Length:1426        Min.   :   2.0   Length:1426        Min.   : 4.00
##  Class :character   1st Qu.: 240.0   Class :character   1st Qu.: 8.00
##  Mode  :character   Median : 476.5   Mode  :character   Median :10.00
##                     Mean   : 506.0                      Mean   :11.42
##                     3rd Qu.: 748.0                      3rd Qu.:14.00
##                     Max.   :1264.0                      Max.   :41.00
##
##   Word.Order          Category         dependency.length.V1     sub.ID
##  Length:1426        Length:1426        Min.   :-1.793155    Min.   : 1.000
##  Class :character   Class :character   1st Qu.:-0.772753    1st Qu.: 1.000
##  Mode  :character   Mode  :character   Median :-0.156026    Median : 3.000
##                                        Mean   : 0.000000    Mean   : 3.988
##                                        3rd Qu.: 0.605226    3rd Qu.: 5.000
##                                        Max.   : 4.046209    Max.   :27.000
##
##    subject          sub_subset_words   sub_subset_length sub_subset_char_length
```

```
##  Length:1426        Length:1426        Min.   : 1.000    Min.   : 1.000
##  Class :character   Class :character   1st Qu.: 1.000    1st Qu.: 3.000
##  Mode  :character   Mode  :character   Median : 1.000    Median : 4.000
##                                        Mean   : 1.321    Mean   : 4.904
##                                        3rd Qu.: 1.750    3rd Qu.: 6.000
##                                        Max.   :10.000    Max.   :52.000
##
##  sub_difluency sub_hesitation   sub_repair     obj.ID          object
##  Min.   :0     Min.   :0      Min.   :0    Min.   : 1.000   Length:1426
##  1st Qu.:0     1st Qu.:0      1st Qu.:0    1st Qu.: 3.000   Class :character
##  Median :0     Median :0      Median :0    Median : 4.000   Mode  :character
##  Mean   :0     Mean   :0      Mean   :0    Mean   : 5.698
##  3rd Qu.:0     3rd Qu.:0      3rd Qu.:0    3rd Qu.: 7.000
##  Max.   :0     Max.   :0      Max.   :0    Max.   :38.000
##
##  obj_subset_words   obj_subset.length obj_subset_char_length obj_difluency
##  Length:1426        Min.   : 1.000    Min.   : 1.000         Min.   :0
##  Class :character   1st Qu.: 1.000    1st Qu.: 4.000         1st Qu.:0
##  Mode  :character   Median : 1.000    Median : 6.000         Median :0
##                     Mean   : 1.773    Mean   : 7.475         Mean   :0
##                     3rd Qu.: 2.000    3rd Qu.: 9.000         3rd Qu.:0
##                     Max.   :12.000    Max.   :59.000         Max.   :0
##
##  obj_hesitation   obj_repair    iobj.ID       iobject
##  Min.   :0      Min.   :0    Min.   : NA    Length:1426
##  1st Qu.:0      1st Qu.:0    1st Qu.: NA    Class :character
##  Median :0      Median :0    Median : NA    Mode  :character
##  Mean   :0      Mean   :0    Mean   :NaN
##  3rd Qu.:0      3rd Qu.:0    3rd Qu.: NA
##  Max.   :0      Max.   :0    Max.   : NA
##                              NA's   :1426
##  iobj_subset_words  iobj_subset_length iobj_subset_char_length
##  Length:1426        Min.   : NA        Min.   : NA
##  Class :character   1st Qu.: NA        1st Qu.: NA
##  Mode  :character   Median : NA        Median : NA
##                     Mean   :NaN        Mean   :NaN
##                     3rd Qu.: NA        3rd Qu.: NA
##                     Max.   : NA        Max.   : NA
##                     NA's   :1426       NA's   :1426
##  File_unique        length_condition     Order            S_Long
##  Length:1426        Equal :678       Min.   :0.0000   Min.   :-0.3333
##  Class :character   O_Long:536       1st Qu.:0.0000   1st Qu.:-0.3333
##  Mode  :character   S_Long:212       Median :0.0000   Median :-0.3333
##                                      Mean   :0.1921   Mean   :-0.1847
##                                      3rd Qu.:0.0000   3rd Qu.:-0.3333
##                                      Max.   :1.0000   Max.   : 0.6667
##
##      O_Long
##  Min.   :-0.33333
##  1st Qu.:-0.33333
##  Median :-0.33333
##  Mean   : 0.04254
##  3rd Qu.: 0.66667
##  Max.   : 0.66667
```

##

3. Fitting Generalized linear model

```
m1<-glm(Order~S_Long+O_Long,
        data=Transitive_Data,
        family="binomial"
             )

summary(m1)
```

```
##
## Call:
## glm(formula = Order ~ S_Long + O_Long, family = "binomial", data = Transitive_Data)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -0.7288  -0.7288  -0.6356  -0.5010   2.0677
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept) -1.56647    0.08541 -18.341   <2e-16 ***
## S_Long      -0.51535    0.23498  -2.193   0.0283 *
## O_Long       0.30660    0.14249   2.152   0.0314 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 1395.5  on 1425  degrees of freedom
## Residual deviance: 1381.2  on 1423  degrees of freedom
## AIC: 1387.2
##
## Number of Fisher Scoring iterations: 4
```

```
m2<-glm(Order~(S_Long+O_Long)*dependency.length,
        data=Transitive_Data,
        family="binomial"
)

summary(m2)
```

```
##
## Call:
## glm(formula = Order ~ (S_Long + O_Long) * dependency.length,
##     family = "binomial", data = Transitive_Data)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -1.1185  -0.6681  -0.6236  -0.3859   2.5146
##
## Coefficients:
```

```
##                            Estimate Std. Error z value Pr(>|z|)
## (Intercept)                -1.59416    0.09265 -17.206  < 2e-16 ***
## S_Long                     -0.61116    0.25797  -2.369  0.01783 *
## O_Long                      0.36050    0.14459   2.493  0.01266 *
## dependency.length           0.06461    0.08255   0.783  0.43382
## S_Long:dependency.length    0.80843    0.21881   3.695  0.00022 ***
## O_Long:dependency.length -0.32350    0.15603  -2.073  0.03814 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 1395.5  on 1425  degrees of freedom
## Residual deviance: 1351.8  on 1420  degrees of freedom
## AIC: 1363.8
##
## Number of Fisher Scoring iterations: 5
```

b. Written Text Data

1. Load the Data

```
## 'data.frame':    4378 obs. of  28 variables:
##  $ File                   : chr  "../UD_Hindi-HDTB\\hi_hdtb-ud-dev.conllu" "../UD_Hindi-HDTB\\hi_hdtb
##  $ Sent_ID                : int  1 53 55 69 75 83 110 114 117 119 ...
##  $ Sentence               : chr  "                                    483                              "
##  $ Length                 : int  24 12 15 13 15 16 37 13 30 27 ...
##  $ Word.Order             : chr  "OSV" "SOV" "SOV" "SOV" ...
##  $ Category               : chr  "Transitive" "Transitive" "Transitive" "Transitive" ...
##  $ dependency.length      : num  3.32 2.1 3.08 2.36 2.92 ...
##  $ sub.ID                 : int  16 4 1 8 12 12 17 1 1 12 ...
##  $ subject                : chr  "  " "  " "  " "  " "  " ...
##  $ sub_subset_words       : chr  "['    ']" "['   ', '     ', '   ', ' ']" "['  ']" "[' \\u200d  ', '    '
##  $ sub_subset_length      : int  1 4 1 4 4 1 2 1 1 1 ...
##  $ sub_subset_char_length : int  7 18 3 22 22 2 9 8 8 2 ...
##  $ sub_difluency          : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ sub_hesitation         : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ sub_repair             : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ obj.ID                 : int  11 6 12 11 13 6 18 10 2 15 ...
##  $ object                 : chr  "    " "    " "    " "  " ...
##  $ obj_subset_words       : chr  "['     ', '  ', '    ', '    ', '    ', '      ']" "['      ']" "['   ',
##  $ obj_subset.length      : int  6 1 2 3 1 5 1 8 21 2 ...
##  $ obj_subset_char_length : int  36 10 10 19 5 29 6 29 104 10 ...
##  $ obj_difluency          : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ obj_hesitation         : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ obj_repair             : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ iobj.ID                : int  NA NA NA NA NA NA NA NA NA NA ...
##  $ iobject                : chr  "" "" "" "" ...
##  $ iobj_subset_words      : chr  "" "" "" "" ...
##  $ iobj_subset_length     : int  NA NA NA NA NA NA NA NA NA NA ...
##  $ iobj_subset_char_length: int  NA NA NA NA NA NA NA NA NA NA ...
```

2. Processing the Data

```r
word_order_data_txt$File_unique<-paste(word_order_data_txt$File, word_order_data_txt$Sent_ID, sep="")

word_order_data_txt$length_condition<-
  ifelse(word_order_data_txt$sub_subset_length>word_order_data_txt$obj_subset.length,'S_Long',
         ifelse(word_order_data_txt$sub_subset_length<word_order_data_txt$obj_subset.length,'O_Long','E


#Subset Transitive Data
Transitive_Data_txt <- subset(word_order_data_txt, word_order_data_txt$Category=='Transitive')

#Order: SOV=0 OSV=1
Transitive_Data_txt$Order<- ifelse(Transitive_Data_txt$Word.Order=="SOV",0,1)

# Treatment Contrast
#-------------------
# S-Long   O-Long
# ---------------
#   1         0
#   0         1
#  -1        -1
#-------------------

#Contrast matrix:
#  OLong SLong
#Equal -1/3  -1/3
#OL     2/3  -1/3
#SL    -1/3   2/3

Transitive_Data_txt$S_Long<-ifelse(Transitive_Data_txt$length_condition=="S_Long",2/3,-1/3)
Transitive_Data_txt$O_Long<-ifelse(Transitive_Data_txt$length_condition=="O_Long",2/3,-1/3)

Transitive_Data_txt$dependency.length<- scale(Transitive_Data_txt$dependency.length)

summary(Transitive_Data_txt)
```

```
##      File              Sent_ID        Sentence            Length
##  Length:3957        Min.   :    1   Length:3957        Min.   : 4.00
##  Class :character   1st Qu.: 1488   Class :character   1st Qu.:15.00
##  Mode  :character   Median : 5282   Mode  :character   Median :21.00
##                     Mean   : 5643                      Mean   :22.36
##                     3rd Qu.: 9212                      3rd Qu.:28.00
##                     Max.   :13306                      Max.   :67.00
##
##   Word.Order          Category         dependency.length.V1     sub.ID
##  Length:3957        Length:3957        Min.   :-2.389720    Min.   : 1.000
##  Class :character   Class :character   1st Qu.:-0.717707    1st Qu.: 3.000
##  Mode  :character   Mode  :character   Median :-0.078973    Median : 7.000
##                                        Mean   : 0.000000    Mean   : 9.365
##                                        3rd Qu.: 0.614779    3rd Qu.:14.000
##                                        Max.   : 4.586144    Max.   :71.000
##
##    subject          sub_subset_words   sub_subset_length sub_subset_char_length
##  Length:3957        Length:3957        Min.   : 1.000    Min.   :  2.0
```

```
##  Class :character   Class :character   1st Qu.: 1.000    1st Qu.:  6.0
##  Mode  :character   Mode  :character   Median : 1.000    Median :  9.0
##                                        Mean   : 2.371    Mean   : 13.8
##                                        3rd Qu.: 3.000    3rd Qu.: 16.0
##                                        Max.   :32.000    Max.   :165.0
##
##  sub_difluency sub_hesitation   sub_repair     obj.ID         object
##  Min.   :0     Min.   :0      Min.   :0    Min.   : 1.00   Length:3957
##  1st Qu.:0     1st Qu.:0      1st Qu.:0    1st Qu.: 9.00   Class :character
##  Median :0     Median :0      Median :0    Median :14.00   Mode  :character
##  Mean   :0     Mean   :0      Mean   :0    Mean   :15.38
##  3rd Qu.:0     3rd Qu.:0      3rd Qu.:0    3rd Qu.:21.00
##  Max.   :0     Max.   :0      Max.   :0    Max.   :72.00
##
##  obj_subset_words   obj_subset.length obj_subset_char_length obj_difluency
##  Length:3957        Min.   : 1.000    Min.   :  2.00         Min.   :0
##  Class :character   1st Qu.: 1.000    1st Qu.:  6.00         1st Qu.:0
##  Mode  :character   Median : 2.000    Median : 11.00         Median :0
##                     Mean   : 3.194    Mean   : 16.87         Mean   :0
##                     3rd Qu.: 4.000    3rd Qu.: 21.00         3rd Qu.:0
##                     Max.   :37.000    Max.   :163.00         Max.   :0
##
##  obj_hesitation   obj_repair    iobj.ID       iobject
##  Min.   :0      Min.   :0    Min.   : NA    Length:3957
##  1st Qu.:0      1st Qu.:0    1st Qu.: NA    Class :character
##  Median :0      Median :0    Median : NA    Mode  :character
##  Mean   :0      Mean   :0    Mean   :NaN
##  3rd Qu.:0      3rd Qu.:0    3rd Qu.: NA
##  Max.   :0      Max.   :0    Max.   : NA
##                              NA's   :3957
##  iobj_subset_words  iobj_subset_length iobj_subset_char_length
##  Length:3957        Min.   : NA        Min.   : NA
##  Class :character   1st Qu.: NA        1st Qu.: NA
##  Mode  :character   Median : NA        Median : NA
##                     Mean   :NaN        Mean   :NaN
##                     3rd Qu.: NA        3rd Qu.: NA
##                     Max.   : NA        Max.   : NA
##                     NA's   :3957       NA's   :3957
##  File_unique        length_condition       Order             S_Long
##  Length:3957        Length:3957        Min.   :0.00000   Min.   :-0.33333
##  Class :character   Class :character   1st Qu.:0.00000   1st Qu.:-0.33333
##  Mode  :character   Mode  :character   Median :0.00000   Median :-0.33333
##                                        Mean   :0.07152   Mean   :-0.05686
##                                        3rd Qu.:0.00000   3rd Qu.: 0.66667
##                                        Max.   :1.00000   Max.   : 0.66667
##
##      O_Long
##  Min.   :-0.3333
##  1st Qu.:-0.3333
##  Median : 0.6667
##  Mean   : 0.1736
##  3rd Qu.: 0.6667
##  Max.   : 0.6667
##
```

3. Fitting Generalized linear model

```
m1_txt<-glm(Order~S_Long+O_Long,
        data=Transitive_Data_txt,
        family="binomial"
)

summary(m1_txt)
```

```
##
## Call:
## glm(formula = Order ~ S_Long + O_Long, family = "binomial", data = Transitive_Data_txt)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -0.4035  -0.3943  -0.3943  -0.3528   2.3698
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept) -2.57620    0.06627 -38.877   <2e-16 ***
## S_Long      -0.27838    0.17976  -1.549    0.121
## O_Long      -0.04820    0.15296  -0.315    0.753
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 2038.3  on 3956  degrees of freedom
## Residual deviance: 2035.2  on 3954  degrees of freedom
## AIC: 2041.2
##
## Number of Fisher Scoring iterations: 5
```

```
m2_txt<-glm(Order~(S_Long+O_Long)*dependency.length,
            data=Transitive_Data_txt,
            family="binomial"
)

summary(m2_txt)
```

```
##
## Call:
## glm(formula = Order ~ (S_Long + O_Long) * dependency.length,
##     family = "binomial", data = Transitive_Data_txt)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -0.7120  -0.4098  -0.3576  -0.3380   2.8085
##
## Coefficients:
##                           Estimate Std. Error z value Pr(>|z|)
## (Intercept)               -2.60540    0.06821 -38.197  < 2e-16 ***
## S_Long                    -0.27747    0.18294  -1.517 0.129333
```

```
## O_Long                    -0.11248    0.15768  -0.713 0.475624
## dependency.length         -0.26652    0.07012  -3.801 0.000144 ***
## S_Long:dependency.length   0.13827    0.18477   0.748 0.454253
## O_Long:dependency.length -0.42563    0.16799  -2.534 0.011288 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 2038.3  on 3956  degrees of freedom
## Residual deviance: 1994.2  on 3951  degrees of freedom
## AIC: 2006.2
##
## Number of Fisher Scoring iterations: 5
```

Adjuncts

a. Dialogue Data

1. Load the Data

```
## 'data.frame':    1813 obs. of  18 variables:
##  $ File                   : chr  "../../Dialouge Corpus Filttered\\parse_gold_filttered\\Phase1\\Phas
##  $ Sent_ID                : int  5 25 69 71 121 127 132 137 206 224 ...
##  $ Sentence               : chr  "                  " "            " "[incomprehensible]
##  $ Length                 : int  8 6 7 7 5 12 8 10 11 10 ...
##  $ Category               : chr  "obl_vs" "s_obl_v" "s_obl_v" "s_obl_v" ...
##  $ Average.Dependency     : num  1.5 2.5 1.8 1.6 1.67 ...
##  $ sub.ID                 : int  7 1 1 1 3 9 2 2 1 1 ...
##  $ subject                : chr  "  " " " " "[incomprehensible]" " " ...
##  $ sub_subset_words       : chr  "['  ', ' ']" "[' ']" "['[incomprehensible]']" "[' ', ' ']" ...
##  $ sub_subset_length      : int  2 1 1 2 2 1 1 1 1 2 ...
##  $ sub_subset_char_length: int  5 2 18 4 5 3 3 4 2 4 ...
##  $ sub_direction          : chr  "RL" "No Direction" "No Direction" "LR" ...
##  $ obl.ID                 : int  2 2 2 3 1 11 6 9 3 7 ...
##  $ obl                    : chr  "  " " " " " " " " " ...
##  $ obl_subset_words       : chr  "['    ', '    ']" "['   ', '    ']" "[' ']" "['  ']" ...
##  $ obl_subset_length      : int  2 2 1 1 1 1 1 1 3 2 ...
##  $ obl_subset_char_length: int  15 9 2 4 3 4 9 4 17 4 ...
##  $ obl_direction          : chr  "Both" "LR" "No Direction" "No Direction" ...
```

2. Processing the Data

```
Transitive_Data_AdjSV<- subset(word_order_data, word_order_data$Category %in% c('obl_sv','s_obl_v'))

head(Transitive_Data_AdjSV)
```

```
##                                                                                                File
## 2  ../../Dialouge Corpus Filttered\\parse_gold_filttered\\Phase1\\Phase1_gold\\hi_1385_gold.conllu
## 3  ../../Dialouge Corpus Filttered\\parse_gold_filttered\\Phase1\\Phase1_gold\\hi_1385_gold.conllu
## 4  ../../Dialouge Corpus Filttered\\parse_gold_filttered\\Phase1\\Phase1_gold\\hi_1385_gold.conllu
```

```
## 5  ../../Dialouge Corpus Filttered\\parse_gold_filttered\\Phase1\\Phase1_gold\\hi_1385_gold.conllu
## 9  ../../Dialouge Corpus Filttered\\parse_gold_filttered\\Phase1\\Phase1_gold\\hi_1385_gold.conllu
## 11 ../../Dialouge Corpus Filttered\\parse_gold_filttered\\Phase1\\Phase1_gold\\hi_1385_gold.conllu
##    Sent_ID                                Sentence Length Category
## 2       25                                        6  s_obl_v
## 3       69 [incomprehensible]                     7  s_obl_v
## 4       71                                        7  s_obl_v
## 5      121                          10            5   obl_sv
## 9      206                                  11  s_obl_v
## 11     229                        31              8  s_obl_v
##    Average.Dependency sub.ID           subject       sub_subset_words
## 2            2.500000      1                                    [' ']
## 3            1.800000      1 [incomprehensible] ['[incomprehensible]']
## 4            1.600000      1                                 [' ', ' ']
## 5            1.666667      3                               [' ', '10']
## 9            2.555556      1                                    [' ']
## 11           1.833333      1                                 [' ', ' ']
##    sub_subset_length sub_subset_char_length sub_direction obl.ID  obl
## 2                  1                      2  No Direction      2
## 3                  1                     18  No Direction      2
## 4                  2                      4            LR      3
## 5                  2                      5            RL      1
## 9                  1                      2  No Direction      3
## 11                 2                      5            LR      3   31
##                obl_subset_words obl_subset_length obl_subset_char_length
## 2             ['   ', '    ']                  2                      9
## 3                         [' ']                 1                      2
## 4                        [' ']                  1                      4
## 5                        [' ']                  1                      3
## 9  ['   ', '    ', '    ']                    3                     17
## 11                     ['31 ']                  1                      4
##    obl_direction
## 2             LR
## 3   No Direction
## 4   No Direction
## 5   No Direction
## 9             LR
## 11            LR
```

```r
#Three conditions for subj obj length
Transitive_Data_AdjSV$length_condition<-
  ifelse(Transitive_Data_AdjSV$sub_subset_length>Transitive_Data_AdjSV$obl_subset_length,'S_Long',
         ifelse(Transitive_Data_AdjSV$sub_subset_length<Transitive_Data_AdjSV$obl_subset_length,'Adj_Lo

Transitive_Data_AdjSV$Order1<- ifelse(Transitive_Data_AdjSV$Category=="s_obl_v",0,1)
Transitive_Data_AdjSV$Order2<- ifelse(Transitive_Data_AdjSV$Category=="obl_sv",0,1)

Transitive_Data_AdjSV$S_Long<-ifelse(Transitive_Data_AdjSV$length_condition=="S_Long",2/3,-1/3)
Transitive_Data_AdjSV$Adj_Long<-ifelse(Transitive_Data_AdjSV$length_condition=="Adj_Long",2/3,-1/3)

Transitive_Data_AdjSV$Average.Dependency <- scale(Transitive_Data_AdjSV$Average.Dependency)

summary(Transitive_Data_AdjSV)
```

```
##      File              Sent_ID         Sentence             Length
##  Length:1343        Min.   :   3.0   Length:1343        Min.   : 4.00
##  Class :character   1st Qu.: 260.0   Class :character   1st Qu.: 8.00
##  Mode  :character   Median : 493.0   Mode  :character   Median : 9.00
##                     Mean   : 520.7                      Mean   :10.97
##                     3rd Qu.: 770.5                      3rd Qu.:13.00
##                     Max.   :1269.0                      Max.   :39.00
##    Category         Average.Dependency.V1     sub.ID         subject
##  Length:1343        Min.   :-3.894625     Min.   : 1.000   Length:1343
##  Class :character   1st Qu.:-0.779474     1st Qu.: 2.000   Class :character
##  Mode  :character   Median :-0.156444     Median : 3.000   Mode  :character
##                     Mean   : 0.000000     Mean   : 4.465
##                     3rd Qu.: 0.565705     3rd Qu.: 6.000
##                     Max.   : 5.167633     Max.   :31.000
##  sub_subset_words   sub_subset_length sub_subset_char_length sub_direction
##  Length:1343        Min.   : 1.000    Min.   : 1.000         Length:1343
##  Class :character   1st Qu.: 1.000    1st Qu.: 3.000         Class :character
##  Mode  :character   Median : 1.000    Median : 4.000         Mode  :character
##                     Mean   : 1.542    Mean   : 5.621
##                     3rd Qu.: 2.000    3rd Qu.: 7.000
##                     Max.   :10.000    Max.   :38.000
##      obl.ID           obl             obl_subset_words   obl_subset_length
##  Min.   : 1.000   Length:1343        Length:1343        Min.   : 1.00
##  1st Qu.: 2.000   Class :character   Class :character   1st Qu.: 1.00
##  Median : 3.000   Mode  :character   Mode  :character   Median : 1.00
##  Mean   : 4.299                                        Mean   : 1.71
##  3rd Qu.: 5.000                                        3rd Qu.: 2.00
##  Max.   :28.000                                        Max.   :20.00
##  obl_subset_char_length obl_direction      length_condition       Order1
##  Min.   : 1.000         Length:1343        Length:1343        Min.   :0.0000
##  1st Qu.: 4.000         Class :character   Class :character   1st Qu.:0.0000
##  Median : 6.000         Mode  :character   Mode  :character   Median :0.0000
##  Mean   : 7.716                                               Mean   :0.4795
##  3rd Qu.:10.000                                               3rd Qu.:1.0000
##  Max.   :72.000                                               Max.   :1.0000
##      Order2           S_Long            Adj_Long
##  Min.   :0.0000   Min.   :-0.33333   Min.   :-0.33333
##  1st Qu.:0.0000   1st Qu.:-0.33333   1st Qu.:-0.33333
##  Median :1.0000   Median :-0.33333   Median :-0.33333
##  Mean   :0.5205   Mean   :-0.08613   Mean   :-0.03475
##  3rd Qu.:1.0000   3rd Qu.:-0.33333   3rd Qu.: 0.66667
##  Max.   :1.0000   Max.   : 0.66667   Max.   : 0.66667
```

3. Fitting Generalized linear model

```
m1<-glm(Order1~S_Long+Adj_Long,
        data=Transitive_Data_AdjSV,
        family="binomial"
)

summary(m1)
```

```
##
```

```
## Call:
## glm(formula = Order1 ~ S_Long + Adj_Long, family = "binomial",
##     data = Transitive_Data_AdjSV)
##
## Deviance Residuals:
##    Min      1Q  Median      3Q     Max
## -1.201  -1.167  -1.092   1.188   1.265
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept) -0.05773    0.05643  -1.023   0.3062
## S_Long       0.17989    0.13666   1.316   0.1881
## Adj_Long     0.25886    0.12887   2.009   0.0446 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 1859.5  on 1342  degrees of freedom
## Residual deviance: 1855.1  on 1340  degrees of freedom
## AIC: 1861.1
##
## Number of Fisher Scoring iterations: 3
```

```
m1_2<-glm(Order1~(S_Long+Adj_Long)*Average.Dependency,
        data=Transitive_Data_AdjSV,
        family="binomial"
)

summary(m1_2)
```

```
##
## Call:
## glm(formula = Order1 ~ (S_Long + Adj_Long) * Average.Dependency,
##     family = "binomial", data = Transitive_Data_AdjSV)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -1.5069  -1.1089  -0.9588   1.2214   1.9172
##
## Coefficients:
##                               Estimate Std. Error z value Pr(>|z|)
## (Intercept)                  -0.01938    0.05756  -0.337  0.73637
## S_Long                        0.18267    0.13750   1.329  0.18400
## Adj_Long                      0.34296    0.13273   2.584  0.00977 **
## Average.Dependency           -0.05600    0.05973  -0.938  0.34850
## S_Long:Average.Dependency     0.08913    0.14351   0.621  0.53454
## Adj_Long:Average.Dependency -0.53952    0.13405  -4.025 5.71e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 1859.5  on 1342  degrees of freedom
```

```
## Residual deviance: 1832.9  on 1337  degrees of freedom
## AIC: 1844.9
##
## Number of Fisher Scoring iterations: 4
```

b. Written Text Data

1. Load the Data

```
## 'data.frame':    3435 obs. of  18 variables:
##  $ File                  : chr  "../../UD_Hindi-HDTB\\hi_hdtb-ud-dev.conllu" "../../UD_Hindi-HDTB\\hi
##  $ Sent_ID               : int  2 6 18 18 19 22 23 24 63 77 ...
##  $ Sentence              : chr  "                                          " "1876
##  $ Length                : int  15 22 34 34 17 15 15 10 33 15 ...
##  $ Category              : chr  "obl_sv" "obl_sv" "obl_sv" "obl_sv" ...
##  $ Average.Dependency    : num  2.46 3.55 5.06 5.06 2.13 ...
##  $ sub.ID                : int  13 16 10 19 14 8 2 4 21 4 ...
##  $ subject               : chr  "   " "    " "   " "   " ...
##  $ sub_subset_words      : chr  "['  \\u200d ', ' ', ' \\u200d    ', ' \\u200d ']" "['   \\u200d  \\u20
##  $ sub_subset_length     : int  4 4 2 1 2 3 3 3 1 3 ...
##  $ sub_subset_char_length: int  23 40 11 6 17 19 12 17 2 16 ...
##  $ sub_direction         : chr  "RL" "RL" "RL" "No Direction" ...
##  $ obl.ID                : int  7 12 8 18 11 3 8 6 23 8 ...
##  $ obl                   : chr  "     " " " " "  " "  " ...
##  $ obl_subset_words      : chr  "['       ']" "[' ']" "['  ']" "['  ']" ...
##  $ obl_subset_length     : int  1 1 1 1 1 3 4 2 2 6 ...
##  $ obl_subset_char_length: int  13 2 4 4 4 20 14 12 12 40 ...
##  $ obl_direction         : chr  "LR" "No Direction" "No Direction" "No Direction" ...
```

2. Processing the Data

```
Transitive_Data_AdjSV_txt<- subset(word_order_data_txt, word_order_data_txt$Category %in% c('obl_sv','s

head(Transitive_Data_AdjSV_txt)
```

```
##                                       File Sent_ID
## 1 ../../UD_Hindi-HDTB\\hi_hdtb-ud-dev.conllu       2
## 2 ../../UD_Hindi-HDTB\\hi_hdtb-ud-dev.conllu       6
## 3 ../../UD_Hindi-HDTB\\hi_hdtb-ud-dev.conllu      18
## 4 ../../UD_Hindi-HDTB\\hi_hdtb-ud-dev.conllu      18
## 5 ../../UD_Hindi-HDTB\\hi_hdtb-ud-dev.conllu      19
## 6 ../../UD_Hindi-HDTB\\hi_hdtb-ud-dev.conllu      22
##
## 1
## 2                                       1876                    ,
## 3                 (              )      (                    )
## 4                 (              )      (                    )
## 5                                                                                ,
## 6
##   Length Category Average.Dependency sub.ID subject
## 1     15   obl_sv           2.461538     13
```

```
## 2     22   obl_sv           3.550000     16
## 3     34   obl_sv           5.062500     10
## 4     34   obl_sv           5.062500     19
## 5     17   obl_sv           2.133333     14
## 6     15   obl_sv           1.692308      8
##                                                  sub_subset_words
## 1             [' \\u200d ', ' ', ' \\u200d    ', ' \\u200d ']
## 2 ['   \\u200d  \\u200d ', ' \\u200d   ', '  ', '     ']
## 3                                                ['  ', '  ']
## 4                                                      ['  ']
## 5                                           ['   ', '     ']
## 6                                 ['   ', '   ', ' \\u200d  ']
##   sub_subset_length sub_subset_char_length sub_direction obl.ID       obl
## 1                 4                     23            RL      7
## 2                 4                     40            RL     12
## 3                 2                     11            RL      8
## 4                 1                      6  No Direction     18
## 5                 2                     17            RL     11
## 6                 3                     19            RL      3
##                     obl_subset_words obl_subset_length obl_subset_char_length
## 1                     ['      ']                 1                     13
## 2                            [' ']                 1                      2
## 3                           ['  ']                 1                      4
## 4                           ['  ']                 1                      4
## 5                           ['  ']                 1                      4
## 6 [' ', ' \\u200d   ', '    ']                 3                     20
##   obl_direction
## 1            LR
## 2  No Direction
## 3  No Direction
## 4  No Direction
## 5  No Direction
## 6            RL
```

```
#Three conditions for subj obj length
Transitive_Data_AdjSV_txt$length_condition<-
  ifelse(Transitive_Data_AdjSV_txt$sub_subset_length>Transitive_Data_AdjSV_txt$obl_subset_length,'S_Long
         ifelse(Transitive_Data_AdjSV_txt$sub_subset_length<Transitive_Data_AdjSV_txt$obl_subset_length


Transitive_Data_AdjSV_txt$Order1<- ifelse(Transitive_Data_AdjSV_txt$Category=="s_obl_v",0,1)
Transitive_Data_AdjSV_txt$Order2<- ifelse(Transitive_Data_AdjSV_txt$Category=="obl_sv",0,1)

Transitive_Data_AdjSV_txt$S_Long<-ifelse(Transitive_Data_AdjSV_txt$length_condition=="S_Long",2/3,-1/3)
Transitive_Data_AdjSV_txt$Adj_Long<-ifelse(Transitive_Data_AdjSV_txt$length_condition=="Adj_Long",2/3,-

Transitive_Data_AdjSV_txt$Average.Dependency <- scale(Transitive_Data_AdjSV_txt$Average.Dependency)

summary(Transitive_Data_AdjSV_txt)
```

```
##      File              Sent_ID        Sentence            Length
##  Length:3412        Min.   :    2   Length:3412        Min.   : 5.00
##  Class :character   1st Qu.: 1443   Class :character   1st Qu.:15.00
##  Mode  :character   Median : 5065   Mode  :character   Median :21.00
```

```
##                     Mean   : 5539                      Mean   :21.98
##                     3rd Qu.: 9154                      3rd Qu.:28.00
##                     Max.   :13298                      Max.   :74.00
##    Category         Average.Dependency.V1     sub.ID        subject
##  Length:3412        Min.   :-2.308871     Min.   : 1.00   Length:3412
##  Class :character   1st Qu.:-0.745004     1st Qu.: 4.00   Class :character
##  Mode  :character   Median :-0.048516     Median : 9.00   Mode  :character
##                     Mean   : 0.000000     Mean   :11.01
##                     3rd Qu.: 0.628147     3rd Qu.:16.00
##                     Max.   : 5.924047     Max.   :72.00
##  sub_subset_words   sub_subset_length sub_subset_char_length sub_direction
##  Length:3412        Min.   : 1.000    Min.   :  2.0          Length:3412
##  Class :character   1st Qu.: 1.000    1st Qu.:  6.0          Class :character
##  Mode  :character   Median : 2.000    Median : 11.0          Mode  :character
##                     Mean   : 2.893    Mean   : 15.7
##                     3rd Qu.: 3.000    3rd Qu.: 19.0
##                     Max.   :45.000    Max.   :270.0
##      obl.ID          obl             obl_subset_words   obl_subset_length
##  Min.   : 1.00   Length:3412        Length:3412        Min.   : 1.000
##  1st Qu.: 4.00   Class :character   Class :character   1st Qu.: 1.000
##  Median : 9.00   Mode  :character   Mode  :character   Median : 2.000
##  Mean   :10.69                                         Mean   : 2.968
##  3rd Qu.:15.00                                         3rd Qu.: 4.000
##  Max.   :57.00                                         Max.   :25.000
##  obl_subset_char_length obl_direction      length_condition       Order1
##  Min.   :  1.00         Length:3412        Length:3412        Min.   :0.0000
##  1st Qu.:  7.00         Class :character   Class :character   1st Qu.:0.0000
##  Median : 13.00         Mode  :character   Mode  :character   Median :1.0000
##  Mean   : 17.56                                               Mean   :0.5094
##  3rd Qu.: 23.00                                               3rd Qu.:1.0000
##  Max.   :140.00                                               Max.   :1.0000
##      Order2           S_Long            Adj_Long
##  Min.   :0.0000   Min.   :-0.33333   Min.   :-0.33333
##  1st Qu.:0.0000   1st Qu.:-0.33333   1st Qu.:-0.33333
##  Median :0.0000   Median :-0.33333   Median :-0.33333
##  Mean   :0.4906   Mean   : 0.02687   Mean   : 0.08109
##  3rd Qu.:1.0000   3rd Qu.: 0.66667   3rd Qu.: 0.66667
##  Max.   :1.0000   Max.   : 0.66667   Max.   : 0.66667
```

3. Fitting Generalized linear model

```
m1_txt<-glm(Order1~S_Long+Adj_Long,
            data=Transitive_Data_AdjSV_txt,
            family="binomial"
)

summary(m1_txt)
```

```
##
## Call:
## glm(formula = Order1 ~ S_Long + Adj_Long, family = "binomial",
##     data = Transitive_Data_AdjSV_txt)
##
```

```
## Deviance Residuals:
##    Min      1Q  Median      3Q     Max
## -1.280  -1.161   1.079   1.194   1.217
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept)  0.03488    0.03550   0.983  0.32581
## S_Long       0.27608    0.09222   2.994  0.00275 **
## Adj_Long    -0.05440    0.08966  -0.607  0.54399
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 4728.8  on 3411  degrees of freedom
## Residual deviance: 4709.5  on 3409  degrees of freedom
## AIC: 4715.5
##
## Number of Fisher Scoring iterations: 3
```

```
m1_II_txt <- glm(Order1~(S_Long+Adj_Long)*Average.Dependency,
                 data=Transitive_Data_AdjSV_txt,
                 family = "binomial")

summary(m1_II_txt)
```

```
##
## Call:
## glm(formula = Order1 ~ (S_Long + Adj_Long) * Average.Dependency,
##     family = "binomial", data = Transitive_Data_AdjSV_txt)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -1.6885  -1.1610   0.8684   1.1806   1.7767
##
## Coefficients:
##                               Estimate Std. Error z value Pr(>|z|)
## (Intercept)                    0.05131    0.03582   1.433 0.152001
## S_Long                         0.30727    0.09292   3.307 0.000944 ***
## Adj_Long                      -0.03633    0.09009  -0.403 0.686734
## Average.Dependency            -0.01426    0.03675  -0.388 0.697955
## S_Long:Average.Dependency      0.28439    0.09470   3.003 0.002672 **
## Adj_Long:Average.Dependency -0.32517    0.09316  -3.491 0.000482 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 4728.8  on 3411  degrees of freedom
## Residual deviance: 4650.5  on 3406  degrees of freedom
## AIC: 4662.5
##
## Number of Fisher Scoring iterations: 4
```