Drawing syntax: analyses

Sofia Stroustrup & Mikkel Wallentin

14 December 2016

In this document we include code and output from statistical analyses made in R on the dataset for the manuscript: Stroustrup & Wallentin: What's left is the subject: Syntax influences lateralized imagery for sentences

---

Load data

```
setwd("~/Documents/matlab_files/tegne_saetninger")
library(xlsx)
tegneData <- read.xlsx("drawing_syntax_data.xlsx","new_coding_consensus")
```

Preprocess data

```
# Remove one drawing that did not contain human characters
tegn<-subset(tegneData, Male_drawn_left!="na")

#Make factor out of dependent variable
tegn$Male_drawn_left<-factor(tegn$Male_drawn_left)

#Make a factor out of verbs
tegntemp<-subset(tegn,select = c(Slaar,Sparker,Kysser,Traekker,Forfoelger,Skubber,Udpeger, Skyder ))
tegn$verbs<- factor(apply(tegntemp, 1, function(x) which(x == 1)), labels = colnames(tegntemp))
rm(tegntemp)
```

Descriptive statistics

A chi-squared test to see if males are on average drawn more to the left.

```
#A chi-squared test to see if males are on average drawn more to the left.
prop.test(sum(tegn$Male_drawn_left==1), length(tegn$Male_drawn_left))
```

```
##
##  1-sample proportions test with continuity correction
##
## data:  sum(tegn$Male_drawn_left == 1) out of length(tegn$Male_drawn_left), null probability 0.5
## X-squared = 3.4637, df = 1, p-value = 0.06273
## alternative hypothesis: true p is not equal to 0.5
## 95 percent confidence interval:
##  0.4986471 0.5528879
## sample estimates:
```

```
##         p
## 0.5258427
```

A trend towards males being drawn more often to the left than females, but not significant.

A table showing the proportion of males drawn left given the independent variables

```
# A table showing the proportion of males drawn left as a function of the independent variables
aggdata <-aggregate(as.numeric(tegn$Male_drawn_left)-1,
    by=list(tegn$Male_first, tegn$Male_active,tegn$Male_subject),
                    FUN=mean, na.rm=TRUE)
library(reshape)
names(aggdata)<-c('Male_first','Male_active','Male_subject','% Male left')
aggdata
```

```
##   Male_first Male_active Male_subject % Male left
## 1          0           0            0   0.2994012
## 2          1           0            0   0.4491018
## 3          0           1            0   0.5389222
## 4          1           1            0   0.6566265
## 5          0           0            1   0.3293413
## 6          1           0            1   0.5209581
## 7          0           1            1   0.6467066
## 8          1           1            1   0.7664671
```

Male drawn on left given 0) Female named first, 1) Male first

```
#Male drawn on left given 0) Female first, 1) Male first
aggdataF<-aggregate(as.numeric(tegn$Male_drawn_left)-1,
    by=list(tegn$Male_first),
                    FUN=mean, na.rm=TRUE)
names(aggdataF)<-c('Male_first','% Male left')
aggdataF
```

```
##   Male_first % Male left
## 1          0   0.4535928
## 2          1   0.5982009
```

Male drawn on left given 0) Female Active part, 1) Male active part

```
#Male drawn on left given 0) Female Active part, 1) Male active part
aggdataA<-aggregate(as.numeric(tegn$Male_drawn_left)-1,
    by=list(tegn$Male_active),
                    FUN=mean, na.rm=TRUE)
names(aggdataA)<-c('Male_active','% Male left')
aggdataA
```

```
##   Male_active % Male left
## 1           0   0.3997006
## 2           1   0.6521739
```

Male drawn on left given 0) Female Subject, 1) Male subject

```
#Male drawn on left given 0) Female Subject, 1) Male subject
aggdataS<-aggregate(as.numeric(tegn$Male_drawn_left)-1,
    by=list(tegn$Male_subject),
                    FUN=mean, na.rm=TRUE)
names(aggdataS)<-c('Male_subject','% Male left')
aggdataS
```

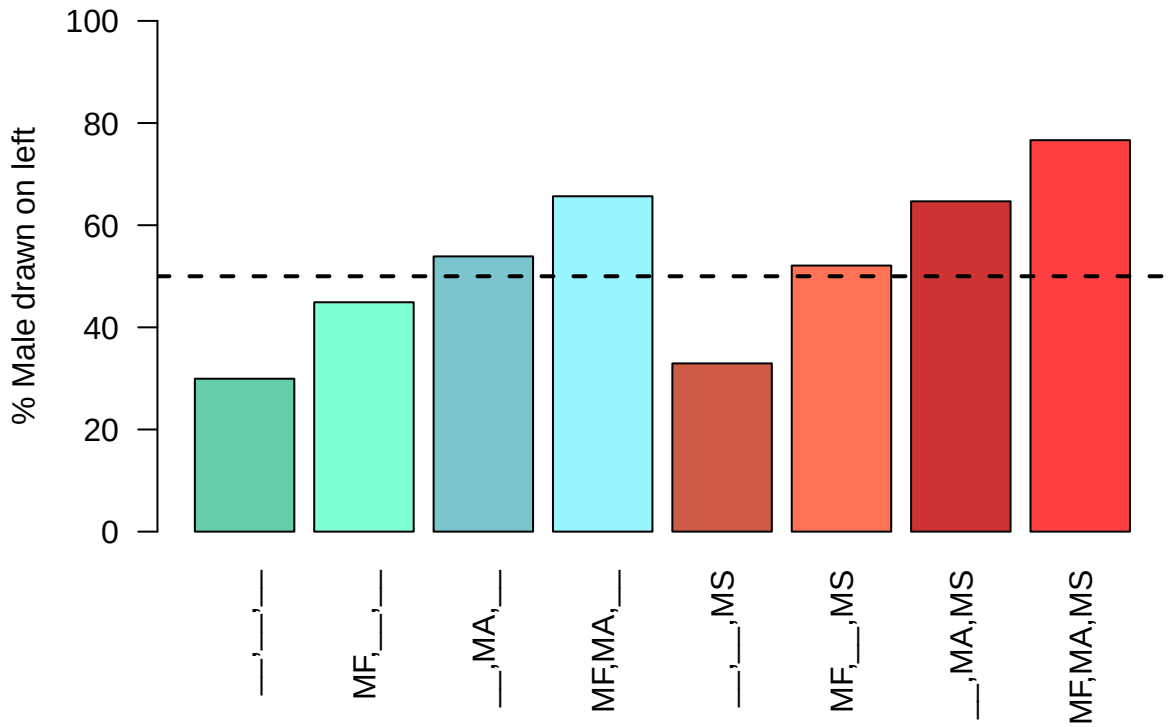
```
##   Male_subject % Male left
## 1            0   0.4857571
## 2            1   0.5658683
```

Bar diagram for all conditions with line indicating 50% chance level

```
aggdata <-aggregate(as.numeric(tegn$Male_drawn_left)-1,
    by=list(tegn$Male_first, tegn$Male_active,tegn$Male_subject),
                    FUN=mean, na.rm=TRUE)

names<-c('__,__,__','MF,__,__','__,MA,__','MF,MA,__','__,__,MS','MF,__,MS','__,MA,MS','MF,MA,MS')

barplot(aggdata$x*100,
        ylim=c(0,100),
        ylab="% Male drawn on left",
        names.arg=names,
        las=2,
        col  = c("aquamarine3","aquamarine","cadetblue3",
                 "cadetblue1","coral3","coral1","brown3","brown1"))
abline(h=50, lty=2, lwd=2)
```



Separate Analysis of verbs

```
library(lme4)
```

```
## Warning: package 'lme4' was built under R version 3.2.5
```

```
#Analysing the effects of verbs
mv <- glmer(Male_drawn_left ~ verbs + (1|Participant)+(1|Stimulus_set), data=tegn,family=binomial)
summary(mv)
```

```
## Generalized linear mixed model fit by maximum likelihood (Laplace
##   Approximation) [glmerMod]
##  Family: binomial  ( logit )
## Formula: Male_drawn_left ~ verbs + (1 | Participant) + (1 | Stimulus_set)
##    Data: tegn
##
##      AIC      BIC   logLik deviance df.resid
##   1860.6   1912.5   -920.3   1840.6     1325
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -1.2445 -1.0411  0.8393  0.9404  1.1691
##
## Random effects:
##  Groups       Name        Variance  Std.Dev.
##  Participant  (Intercept) 4.730e-06 0.002175
##  Stimulus_set (Intercept) 3.293e-02 0.181470
## Number of obs: 1335, groups:  Participant, 167; Stimulus_set, 8
##
## Fixed effects:
##                 Estimate Std. Error z value Pr(>|z|)
## (Intercept)      0.12637    0.16862   0.749    0.454
## verbsSparker     0.04859    0.22043   0.220    0.826
## verbsKysser      0.11131    0.22113   0.503    0.615
## verbsTraekker   -0.14507    0.21999  -0.659    0.510
## verbsForfoelger -0.02423    0.22014  -0.110    0.912
## verbsSkubber     0.02427    0.22032   0.110    0.912
## verbsUdpeger    -0.16922    0.22001  -0.769    0.442
## verbsSkyder     -0.07261    0.22004  -0.330    0.741
##
## Correlation of Fixed Effects:
##             (Intr) vrbsSp vrbsKy vrbsTr vrbsFr vrbsSkb vrbsUd
## verbsSparkr -0.652
## verbsKysser -0.650  0.497
## verbsTrakkr -0.654  0.500  0.498
## verbsFrflgr -0.653  0.500  0.498  0.501
## verbsSkubbr -0.653  0.499  0.498  0.500  0.500
## verbsUdpegr -0.654  0.500  0.498  0.501  0.501  0.500
## verbsSkyder -0.654  0.500  0.498  0.501  0.501  0.500   0.501
```

Model comparison

We compare a complex and a simple model (see main article for details)

```
library(lme4)

#In order to compare models were one has included participant gender,
#we need to exclude one participant who did not reveal his/her gender and
#one who were ambidextrous.

#Participants and stimulus set are included as random effects.

tegn2<-subset(tegn,Participant_Gender !="Not reported")
tegn2<-subset(tegn2,Hand !="A")


#Fitting a simple model with only main effects.
m1 <- glmer(Male_drawn_left ~ Male_first+Male_active+Male_subject
            + (1|Participant)+(1|Stimulus_set), data=tegn2,family=binomial)

#Fitting af complex model including both language bias, handedness,
#gender, main effects and interactions.
m2 <- glmer(Male_drawn_left ~ Male_active*Male_first*Male_subject+Hand+Lang_bias+Participant_Gender+
          (1|Participant)+(1|Stimulus_set),
          data=tegn2,family=binomial)

#Making a model comparison
anova(m1,m2)
```

```
## Data: tegn2
## Models:
## m1: Male_drawn_left ~ Male_first + Male_active + Male_subject + (1 |
## m1:     Participant) + (1 | Stimulus_set)
## m2: Male_drawn_left ~ Male_active * Male_first * Male_subject + Hand +
## m2:     Lang_bias + Participant_Gender + (1 | Participant) + (1 |
## m2:     Stimulus_set)
##    Df    AIC    BIC  logLik deviance  Chisq Chi Df Pr(>Chisq)
## m1  6 1691.6 1722.7 -839.82   1679.6
## m2 13 1701.4 1768.7 -837.68   1675.4 4.2706      7     0.7481
```

We find no differences between the complex and the simple model. Thus we report results from the simple model, using the full data set.

Main analysis using simple model

```
#We re-introduce data from participants who did not reveal their gender
#or were ambidextrous
m3 <- glmer(Male_drawn_left ~ Male_first+Male_active+Male_subject+ (1|Participant)+(1|Stimulus_set),
           data=tegn,family=binomial)

#Finding z-statistics from simple model
summary(m3)
```

```
## Generalized linear mixed model fit by maximum likelihood (Laplace
##   Approximation) [glmerMod]
##  Family: binomial  ( logit )
## Formula: Male_drawn_left ~ Male_first + Male_active + Male_subject + (1 |
##     Participant) + (1 | Stimulus_set)
##    Data: tegn
##
##      AIC      BIC   logLik deviance df.resid
##   1727.6   1758.7   -857.8   1715.6     1329
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -2.0700 -0.8804  0.5231  0.8555  1.8920
##
## Random effects:
##  Groups       Name        Variance Std.Dev.
##  Participant  (Intercept) 0.05936  0.2436
##  Stimulus_set (Intercept) 0.04096  0.2024
## Number of obs: 1335, groups:  Participant, 167; Stimulus_set, 8
##
## Fixed effects:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)   -0.9391     0.1416  -6.631 3.34e-11 ***
## Male_first     0.6452     0.1172   5.505 3.70e-08 ***
## Male_active    1.0935     0.1183   9.240  < 2e-16 ***
## Male_subject   0.3589     0.1166   3.078  0.00208 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##             (Intr) Ml_frs Ml_ctv
## Male_first  -0.457
## Male_active -0.464  0.093
## Male_subjct -0.441  0.029  0.050
```

We find significant effects of Male named first, of Male being activ and of Male being grammatical subject.

Additional control analysis

It was observed that a few participants wrote the sentences down above the drawings. This might have skewed the results.

We therefore make a control analysis in which we exclude the three participants.

```
#Data excluding three additional participants who wrote the sentences above their drawings
tegn3<-subset(tegn,Participant !=c(36))
tegn3<-subset(tegn3,Participant !=c(83))
tegn3<-subset(tegn3,Participant !=c(94))

#Fit the simple model again
m4 <- glmer(Male_drawn_left ~ Male_first + Male_active + Male_subject + (1|Participant)+ (1|Stimulus_se
            data=tegn3,family=binomial)

summary(m4)
```

```
## Generalized linear mixed model fit by maximum likelihood (Laplace
##   Approximation) [glmerMod]
##  Family: binomial  ( logit )
## Formula: Male_drawn_left ~ Male_first + Male_active + Male_subject + (1 |
##     Participant) + (1 | Stimulus_set)
##    Data: tegn3
##
##      AIC      BIC   logLik deviance df.resid
##   1697.1   1728.1   -842.5   1685.1     1305
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -1.9737 -0.8858  0.5144  0.8538  1.8914
##
## Random effects:
##  Groups       Name        Variance Std.Dev.
##  Participant  (Intercept) 0.06722  0.2593
##  Stimulus_set (Intercept) 0.03669  0.1915
## Number of obs: 1311, groups:  Participant, 164; Stimulus_set, 8
##
## Fixed effects:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)   -0.9496     0.1410  -6.734 1.65e-11 ***
## Male_first     0.6639     0.1184   5.608 2.05e-08 ***
## Male_active    1.0827     0.1195   9.062  < 2e-16 ***
## Male_subject   0.3726     0.1177   3.165  0.00155 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##             (Intr) Ml_frs Ml_ctv
## Male_first  -0.465
## Male_active -0.472  0.095
## Male_subjct -0.448  0.031  0.052
```

Results are the same. Thus we report the effects of the analysis which includes the most participants.
