Test link between subject-verb order and interrogative position

Load libraries and data

```
library(lme4)
library(sjPlot)
```

```
d = read.csv("wals-language.csv/language.csv")
```

Make variables

```
d$IP = d$X93A.Position.of.Interrogative.Phrases.in.Content.Questions

d$IP[d$IP %in% c("3 Mixed")] = NA
d$IP[d$IP %in% c("")] = NA
d$IP = as.character(d$IP)
d$IP[d$IP=="1 Initial interrogative phrase"] = "Initial"
d$IP[d$IP=="2 Not initial interrogative phrase"] = "Non-Initial"
d$IP = factor(d$IP)

d$WO = d$X81A.Order.of.Subject..Object.and.Verb

d$SV = d$X82A.Order.of.Subject.and.Verb
d$SV[d$SV==""] = NA
d$SV[d$SV=="3 No dominant order"] = NA
d$SV = d$SV == "1 SV"
```

Contingency table of initial interrogative phrase by basic word order:

```
table(d$WO,d$IP) / rowSums(table(d$WO,d$IP))
```

```
##
##                        Initial Non-Initial
##                      0.1923077   0.8076923
##   1 SOV              0.1790123   0.8209877
##   2 SVO              0.2482759   0.7517241
##   3 VSO              0.7076923   0.2923077
##   4 VOS              0.7692308   0.2307692
##   5 OVS              0.5714286   0.4285714
##   6 OSV              0.0000000   1.0000000
##   7 No dominant order 0.5900000   0.4100000
```

Pattern:

- Subject before verb = non-initial
- Verb before subject = initial

Mixed effects model

Predict order of subject and verb by interrogative phrase position, controlling for family and macroarea.

```
m0 = glmer(IP ~ 1 + (1 + SV | family) + (1 + SV| macroarea),
           data=d[!is.na(d$IP),], family = binomial)
m1 = glmer(IP ~ 1 + SV + (1 + SV| family) + (1 + SV| macroarea),
           data=d[!is.na(d$IP),], family = binomial)
anova(m0,m1)
```

```
## Data: d[!is.na(d$IP), ]
## Models:
## m0: IP ~ 1 + (1 + SV | family) + (1 + SV | macroarea)
## m1: IP ~ 1 + SV + (1 + SV | family) + (1 + SV | macroarea)
##    Df    AIC    BIC  logLik deviance  Chisq Chi Df Pr(>Chisq)
## m0  7 624.36 657.16 -305.18   610.36
## m1  8 610.47 647.95 -297.24   594.47 15.894      1    6.7e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

There was a significant main effect of subject-verb order (beta = 2.6 , std.err = 0.42 , Wald t = 6.2 , Wald p = 6.3e-10 ; log likelihood difference = 7.9 , df = 1 , Chi Squared = 15.89 , p = 6.7e-05).. Probability of initial interrogatives with subject-verb order = 0.2266233. Probability of initial interrogatives with verb-subject = 0.7990444

Look at direction:

```
summary(m1)
```

```
## Generalized linear mixed model fit by maximum likelihood (Laplace
##    Approximation) [glmerMod]
##  Family: binomial  ( logit )
## Formula: IP ~ 1 + SV + (1 + SV | family) + (1 + SV | macroarea)
##    Data: d[!is.na(d$IP), ]
##
##      AIC      BIC   logLik deviance df.resid
##    610.5    647.9   -297.2    594.5      792
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -5.9615 -0.2197  0.1677  0.3607  4.5511
##
## Random effects:
##  Groups    Name        Variance  Std.Dev. Corr
##  family    (Intercept) 9.848e-01 0.992392
##            SVTRUE      6.602e-03 0.081250 1.00
##  macroarea (Intercept) 3.031e+00 1.740881
##            SVTRUE      1.618e-06 0.001272 1.00
## Number of obs: 800, groups:  family, 140; macroarea, 7
##
## Fixed effects:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept)  -1.2275     0.7953  -1.543    0.123
## SVTRUE        2.6078     0.4218   6.182 6.33e-10 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
## 
## Correlation of Fixed Effects:
##       (Intr)
## SVTRUE -0.435
```

Random effects

Note that there are very small differences between random slope for SV by macroarea. The convergence is singulative (total correlation between intercept and slopes), but there are no qualitative differences when removing the random slopes.

```
sjp.glmer(m1, 're',sort.est = "(Intercept)", geom.colors = c(1,1))
```

```
## Plotting random effects...
## Plotting random effects...
```