Testing the relationship between verb position and position of verbal affixes

Introduction

Thompson (1998) points out that interrogative structures make turn transition relevant: a question demands an answer. Thompson argues that, in order to be effective, interrogatives should generally apply to prosodic units, and therefore appear at turn boundaries, rather than in the middle of turns. If interrogatives are morphologically bound to the verb, this constraint leads to a specific prediction: languages that place the verb at the end of a sentence should have interrogative suffixes (so that the interrogative appears after the verb at the boundary), while languages with verbs at the beginning should have interrogative prefixes. In fact, this should bias the language towards suffixing or prefixing in general. We tested this statistically by looking at the probability of suffixes for different word orders in a sample of the world's languages, controlling for historical influence. Indeed, we find that suffixes are much more likely than prefixes in verb-final languages (460 languages taken from Dryer, 2013b and Dryer, 2013a, mixed effects model controlling for language family, log likelihood difference = 12.27, 2 = 24.5, df = 2, $p < 0.0001$, see below).

Load libraries

```
library(lme4)
```

```
## Loading required package: Matrix
```

```
##
## Attaching package: 'lme4'
```

```
## The following object is masked from 'package:stats':
##
##     sigma
```

Load data

```
d.wals = read.csv("wals_data/language.csv",na.string='')
```

Make a variable which marks the position of the verb.

```
OrderSOV = as.character(d.wals$X81A.Order.of.Subject..Object.and.Verb)

d.wals$VerbPos = NA
d.wals[OrderSOV %in% c('1 SOV','6 OSV'),]$VerbPos = 3
d.wals[OrderSOV %in% c('2 SVO','5 OVS'),]$VerbPos = 2
d.wals[OrderSOV %in% c('3 VSO','4 VOS'),]$VerbPos = 1

# Variable for whether language is verb-initial
d.wals$VerbInitial = d.wals$VerbPos==1
```

Make a variable that distinguishes between suffixing and prefixing inflectional morphology.

```
InfMorph = as.character(d.wals$X26A.Prefixing.vs..Suffixing.in.Inflectional.Morphology)

d.wals$Aff = NA
d.wals$Aff[InfMorph=="2 Strongly suffixing"] = 'Suff'
d.wals$Aff[InfMorph=="3 Weakly suffixing"] = 'Suff'
d.wals$Aff[InfMorph=="5 Weakly prefixing"] = 'Pre'
d.wals$Aff[InfMorph=="6 Weakly prefixing"] = 'Pre'
```

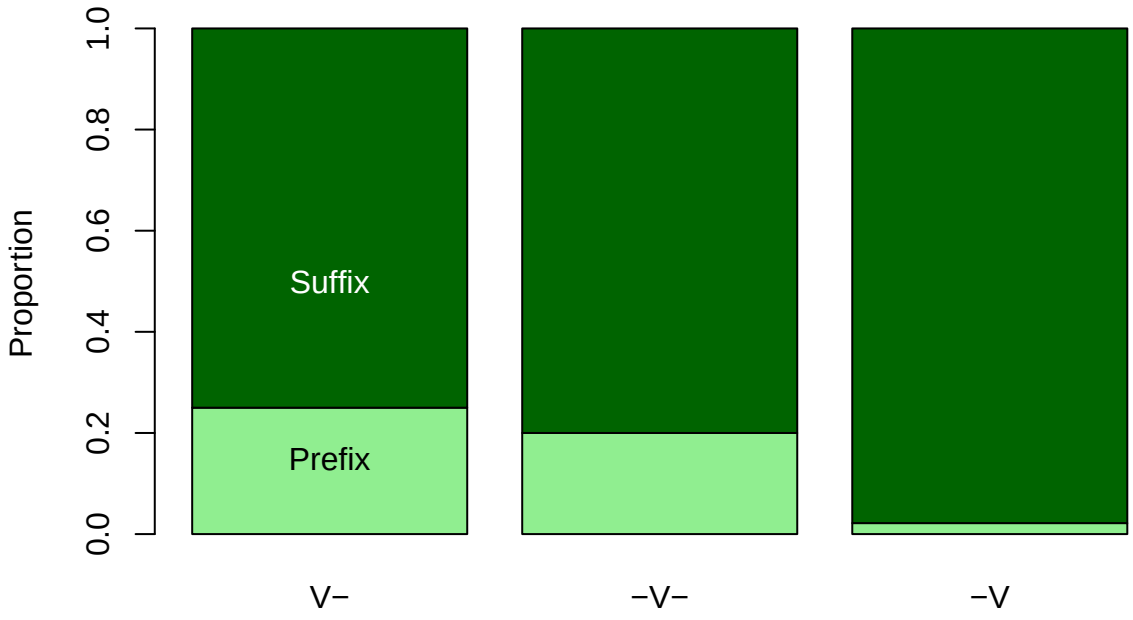
Thompson (1998) predicts that interrogative affixes in v-final languages should be suffixes rather than prefixes. Here are the raw numbers:

```
pqInterMorph = d.wals$X116A.Polar.Questions=='2 Interrogative verb morphology'
table(
  d.wals[pqInterMorph,]$X26A.Prefixing.vs..Suffixing.in.Inflectional.Morphology,
  d.wals[pqInterMorph,]$VerbPos)
```

```
##
##                                   1  2  3
##   1 Little affixation             0  0  0
##   2 Strongly suffixing            1  3 72
##   3 Weakly suffixing              2  1 18
##   4 Equal prefixing and suffixing 2  3  9
##   5 Weakly prefixing              1  1  2
##   6 Strong prefixing              0  0  0
```

Plot the raw numbers

```
tx = table(
  d.wals[pqInterMorph,]$Aff,
  d.wals[pqInterMorph,]$VerbPos)

barplot(
  t(t(tx)/colSums(tx)),
  names.arg=c("V-","-V-","-V"),
  ylab="Proportion",
  col=c("light green","dark green"))
text(0.7,0.5,"Suffix",col="white")
text(0.7,0.15,"Prefix")
```


Proportion
0.0
0.2
0.4
0.6
0.8
1.0
Suffix
Prefix
V−
−V−
−V

Mixed effects model

A mixed effects model, predicing the presence of suffixing by the position of the verb, with a random effect for language family.

```
# Null model, without fixed effect for verb pos.
lm0 = glmer(
  Aff=="Suff"~1+(1|family),
  data=d.wals[!is.na(d.wals$VerbPos),],
  family=binomial(link=logit))

# Main model.
lm1 = glmer(
  Aff=="Suff"~factor(VerbPos)+(1|family),
  data=d.wals[!is.na(d.wals$VerbPos),],
  family=binomial(link=logit))
```

Compare fit of models. Verb position is a significant predictor of the probability of suffixing verb morphology.

```
suffTest = anova(lm0,lm1)
suffTest
```

```
## Data: d.wals[!is.na(d.wals$VerbPos), ]
## Models:
## lm0: Aff == "Suff" ~ 1 + (1 | family)
## lm1: Aff == "Suff" ~ factor(VerbPos) + (1 | family)
##     Df    AIC    BIC  logLik deviance  Chisq Chi Df Pr(>Chisq)
## lm0  2 305.77 314.04 -150.89   301.77
## lm1  4 285.25 301.77 -138.62   277.25 24.523      2   4.73e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Look at model estimates. Note that the strongest difference is between verb-initial and verb-final languages.

```
summary(lm1)
```

```
## Generalized linear mixed model fit by maximum likelihood (Laplace
##   Approximation) [glmerMod]
##  Family: binomial  ( logit )
## Formula: Aff == "Suff" ~ factor(VerbPos) + (1 | family)
##    Data: d.wals[!is.na(d.wals$VerbPos), ]
##
##      AIC      BIC   logLik deviance df.resid
##    285.2    301.8   -138.6    277.2      456
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -6.1603  0.0939  0.1535  0.2355  2.2787
##
## Random effects:
##  Groups Name        Variance Std.Dev.
##  family (Intercept) 3.662    1.914
## Number of obs: 460, groups:  family, 87
##
## Fixed effects:
##                  Estimate Std. Error z value Pr(>|z|)
## (Intercept)        1.0781     0.6825   1.579    0.114
```

```
## factor(VerbPos)2   0.5197     0.5657   0.919    0.358
## factor(VerbPos)3   2.5017     0.6371   3.927 8.62e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##             (Intr) f(VP)2
## fctr(VrbP)2 -0.602
## fctr(VrbP)3 -0.663  0.712
```

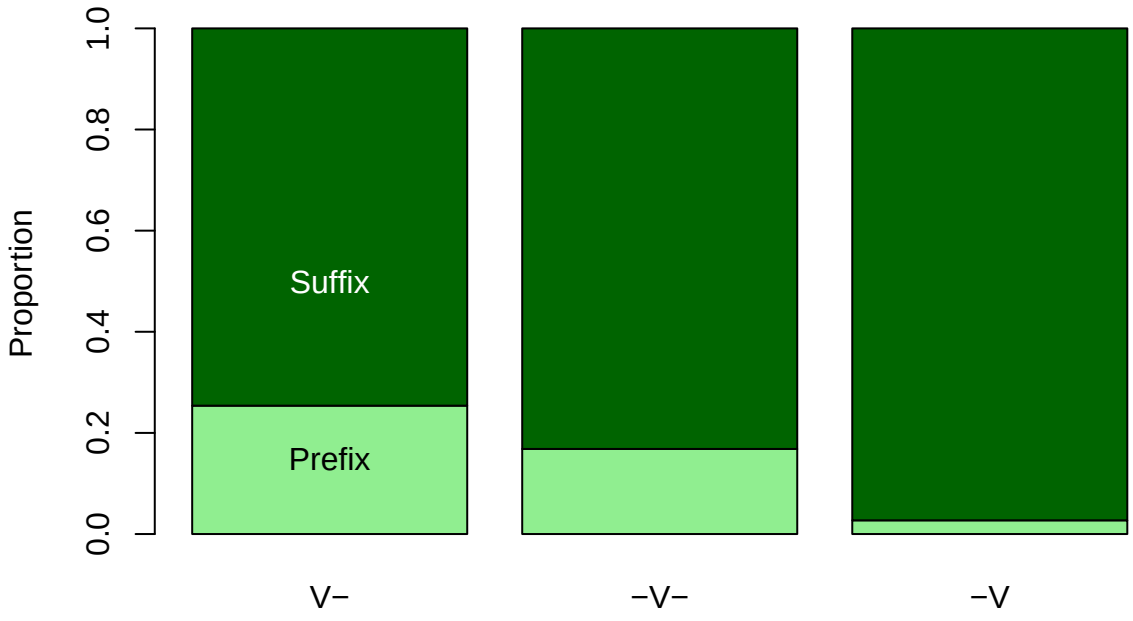
Use fixed effects to estimate probability of suffixing morphology given verb position.

```
est1 = c(fixef(lm1)[1],sum(fixef(lm1)[1:2]),sum(fixef(lm1)[c(1,3)]))
tx2 = exp(est1)/(1+exp(est1))
tx2
```

```
## (Intercept)
##   0.7461282   0.8317084   0.9728741
```

Plot the model estimates.

```
barplot(
  rbind(1-tx2,tx2),
  names.arg=c("V-","-V-","-V"),
  ylab="Proportion",
  col=c("light green","dark green"))
text(0.7,0.5,"Suffix",col="white")
text(0.7,0.15,"Prefix")
```



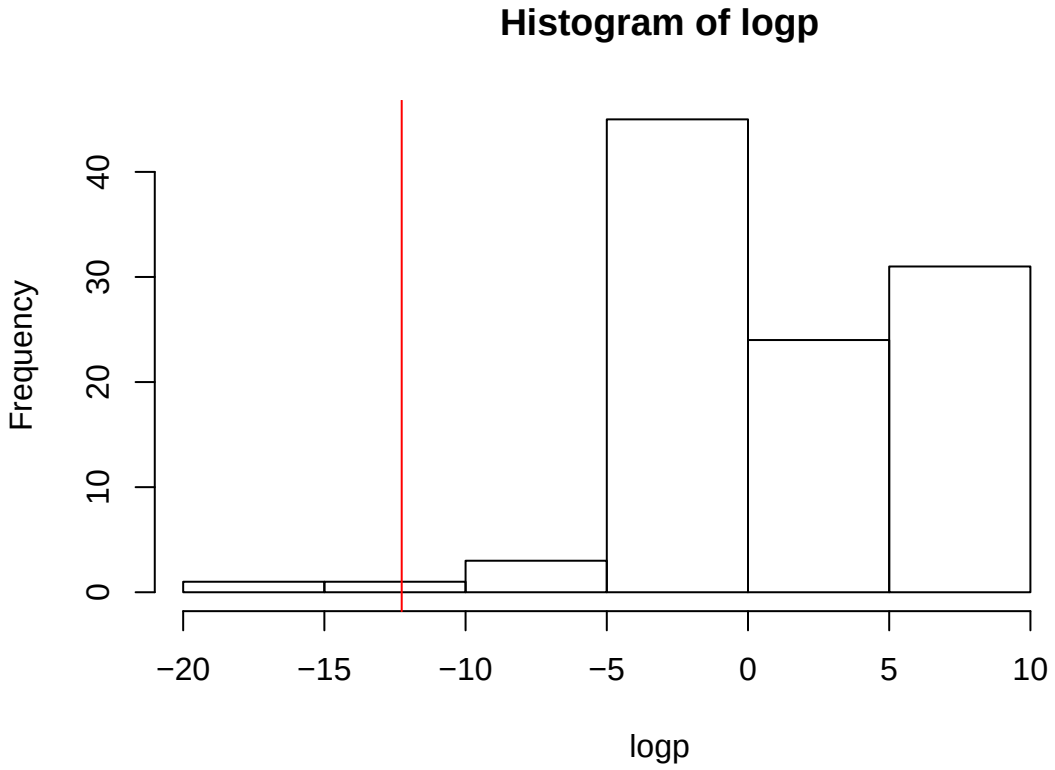
Comparison to other features

Below we test whether the link between suffixing position and word order is stronger than the links between suffixing position and other linguistic variables in WALS. We expect at least 95% of variables to exhibit a weaker link

```
res = data.frame(var='',chi='',df='',p='', n='', stringsAsFactors = F)

for(i in 9:ncol(d.wals)){

  n = sum(!is.na(d.wals[,i]) & !is.na(d.wals$Aff))
  if(n > 50){

  lmA0 =
    tryCatch(
      glmer(
        Aff=="Suff"~1+(1|family),
        data=d.wals[!is.na(d.wals[,i]) & !is.na(d.wals$Aff),],
        family=binomial(link=logit)),
      error=function(e) NA,
      warning=function(e) NA
    )

  lmA1 =
    tryCatch(
        glmer(
      Aff=="Suff"~
        factor(d.wals[!is.na(d.wals[,i]) & !is.na(d.wals$Aff),i])+
        (1|family),
      data=d.wals[!is.na(d.wals[,i]) & !is.na(d.wals$Aff),],
      family=binomial(link=logit)),
      error=function(e) NA,
      warning=function(e) NA
    )

  if(!is.na(lmA1) & !is.na(lmA0)){
    a = anova(lmA0,lmA1)
    res = rbind(res,
                c(colnames(d.wals)[i],
                  a$Chisq[2],
                  a$`Chi Df`[2],
                  a$`Pr(>Chisq)`[2]),
                  n)
  }
  }
}
res$p = as.numeric(res$p)
res$n = as.numeric(res$n)
res = res[!is.na(res$p),]
# Remove word order variables
res = res[!grepl("Order.of", res$var),]
write.csv(res,'SerendipityTest.csv')
```

The models converged for 105 variables.

```
logp = log(res$p)
suffAndWordOrder.p = log(suffTest$`Pr(>Chisq)`[2])

hist(logp, xlim=c(min(min(logp),suffAndWordOrder.p),max(logp)+max(logp)*1.1 ))
abline(v=suffAndWordOrder.p, col=2)
```



Proportion of tests with lower p-value:

```
sum(logp <suffAndWordOrder.p) / length(logp)
```

```
## [1] 0.00952381
```

List of variables with lower p-value:

```
res[res$p < suffTest$`Pr(>Chisq)`[2], ]$var
```

```
## [1] "X57A.Position.of.Pronominal.Possessive.Affixes"
```

This test concludes that the link between affix position and word order, as found in the sections above, is stronger than the vast majority of other possible links.

References

Dryer, M. S. (2013a). Order of Subject, Object and Verb. Max Planck Institute for Evolutionary Anthropology, Leipzig.

Dryer, M. S. (2013b). Prefixing vs. Suffixing in Inflectional Morphology. Max Planck Institute for Evolutionary Anthropology, Leipzig.

Thompson, S. A. (1998). A discourse explanation for the cross-linguistic differences in the grammar of interrogation and negation. Case, typology and grammar: In honor of Barry J. Blake, 309-341.