

# Supporting Information for “Conversation, cognition and cultural evolution: a model of the cultural evolution of word order through pressures imposed from turn taking in conversation”

Seán G. Roberts and Stephen C. Levinson

This document contains supporting information:

- 1) The R code and results for the test of the relationship between verb position and verbal affixes. This was produced using RMarkdown.
- 2) The code and results for estimating the typological distributions of languages while controlling for historical influence. This was produced using RMarkdown.
- 3) A summary of some robustness tests of the model, including the ratio of turns to conversations; populations size and noise; initial conditions; and transitions between states.

A github repository of the data, model code, statistical code and results is available here: <https://github.com/seannyD/TurnTakingCulturalEvolution>.

# 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

```

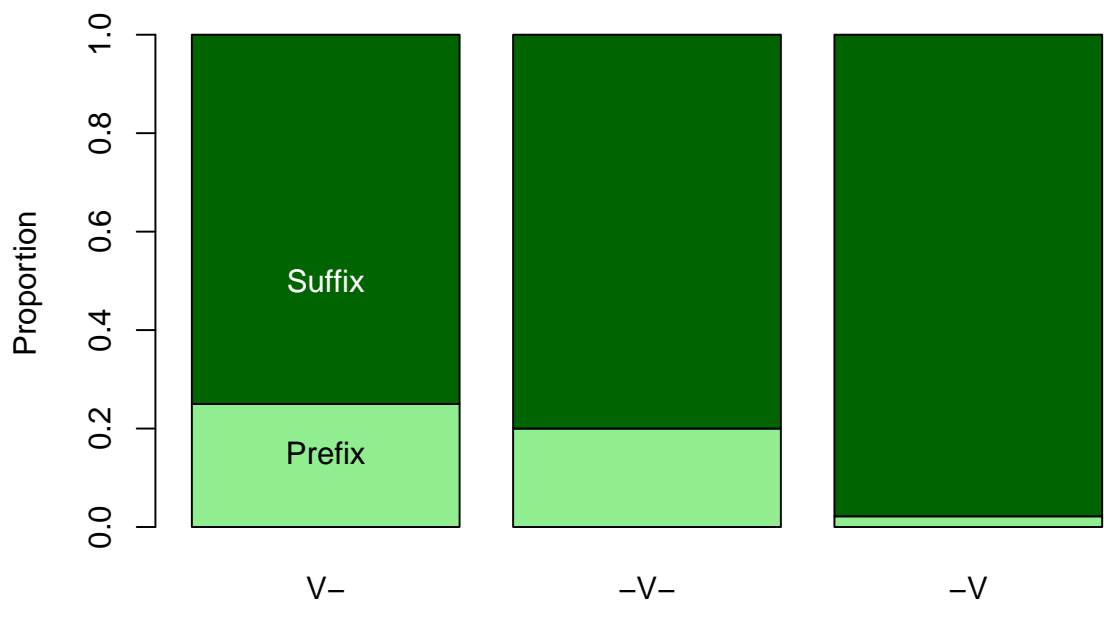
Plot the raw numbers

```

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")

```



## Mixed effects model

A mixed effects model, predicting 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
```

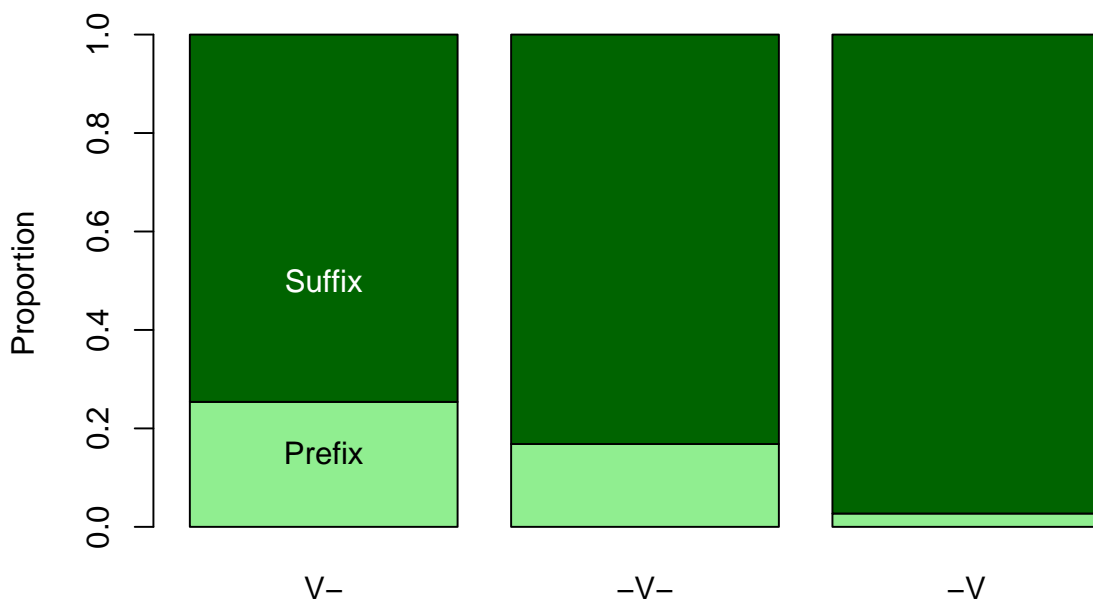
Use fixed effects to estimate probability of suffixing morphology given verb position.

```
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.

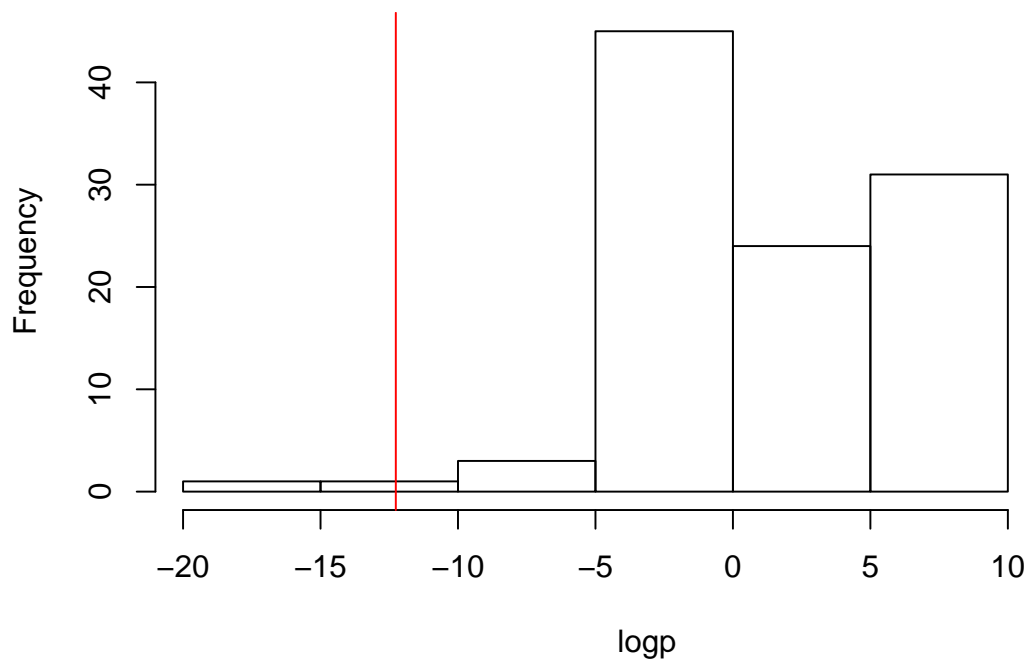
```

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)

```

## Histogram of logp



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.



# Estimating the typological distribution of word order and sentence particles

Load libraries.

```
library(lme4)
```

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

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

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

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
```

Make a centered variable from the verb position.

```
d.wals$VerbPos.norm = d.wals$VerbPos/3
```

Make a variable for the position of particles. Code only initial, final and none, removing 'second position' and 'other position'.

```
# make a shortcut for position of polar question particles
QP = as.character(d.wals$X92A.Position.of.Polar.Question.Particles)

# find non-NA datapoints
QPNA = !is.na(d.wals$X92A.Position.of.Polar.Question.Particles)

d.wals$QP.pos3 = NA
d.wals$QP.pos3[QPNA & QP=='1 Initial'] = "3 Initial"
d.wals$QP.pos3[QPNA & QP=='3 Second position'] = NA
d.wals$QP.pos3[QPNA & QP=='4 Other position'] = NA
d.wals$QP.pos3[QPNA & QP=='2 Final'] = "2 Final"
d.wals$QP.pos3[QP=='6 No question particle'] = "1 None"
```

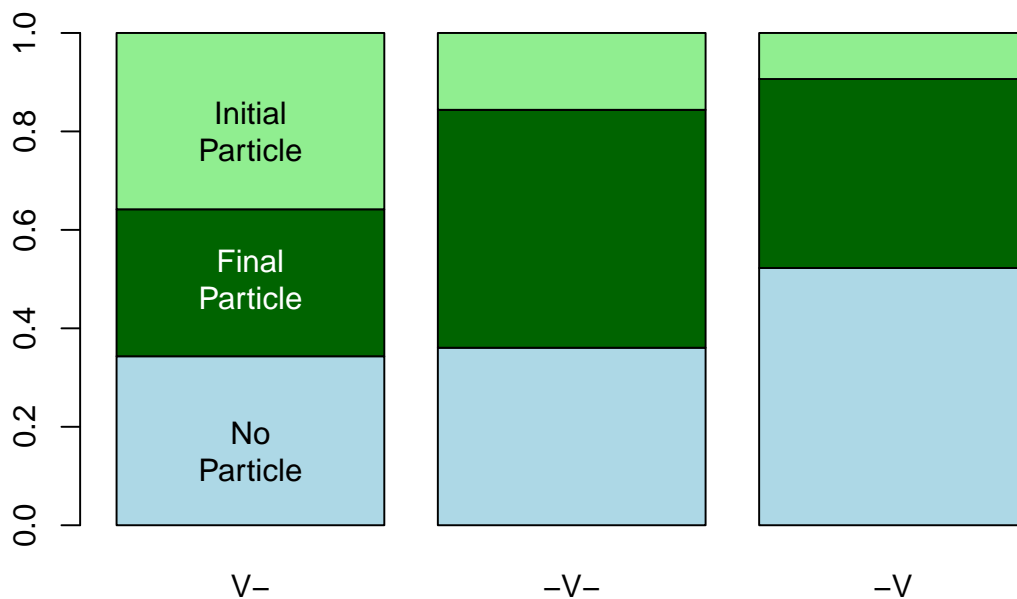
Show the raw data. This is just the number of languages with each type, uncorrected for historical relations.

```
table(d.wals$QP.pos3, d.wals$VerbPos)
```

```
##
##           1  2  3
## 1 None    23 97 162
## 2 Final   20 130 119
## 3 Initial  24  42  29
```

Plot the raw data.

```
#
tx4 = table(d.wals$QP.pos3,d.wals$VerbPos)
barplot(
  t(t(tx4)/colSums(tx4)),
  names.arg=c("V-", "-V-", "-V"),
  col=rev(c("light green","dark green",'light blue')))
text(0.7,0.8,"Initial\nParticle")
text(0.7,0.5,"Final\nParticle",col='white')
text(0.7,0.15,"No\nParticle")
```



## Mixed effects models

Make three mixed effects models, each predicting the probability of a verb being in a given position by the position of the polar question marker. There is a random effect for families, which adds a basic control for historical relatedness.

```
lmVSO = glmer(VerbPos==1~QP.pos3+(1|family),data=d.wals,family=binomial(link=logit))
lmSVO = glmer(VerbPos==2~QP.pos3+(1|family),data=d.wals,family=binomial(link=logit))
lmSOV = glmer(VerbPos==3~QP.pos3+(1|family),data=d.wals,family=binomial(link=logit))
```

Use the fixed effects of the models to estimate the proportion of each question particle type for each basic word order type. This estimate should take into account the historical relations between families.

```

x = c(exp(fixef(lmVSO)[1]),exp(sum(fixef(lmVSO)[1:2])),exp(sum(fixef(lmVSO)[c(1,3)])))
vsoP = x/sum(x)
x = c(exp(fixef(lmSVO)[1]),exp(sum(fixef(lmSVO)[1:2])),exp(sum(fixef(lmSVO)[c(1,3)])))
svoP = x/sum(x)
x = c(exp(fixef(lmSOV)[1]),exp(sum(fixef(lmSOV)[1:2])),exp(sum(fixef(lmSOV)[c(1,3)])))
sovP = x/sum(x)

```

Show the model estimates.

```

estX = rbind(vsoP,svoP,sovP)
estX = estX/rowSums(estX)
colnames(estX) = c("None","Final","Initial")
estX

```

```

##           None      Final      Initial
## vsoP 0.1470058 0.2144506 0.6385436
## svoP 0.2875475 0.4135087 0.2989438
## sovP 0.5150098 0.2746864 0.2103037

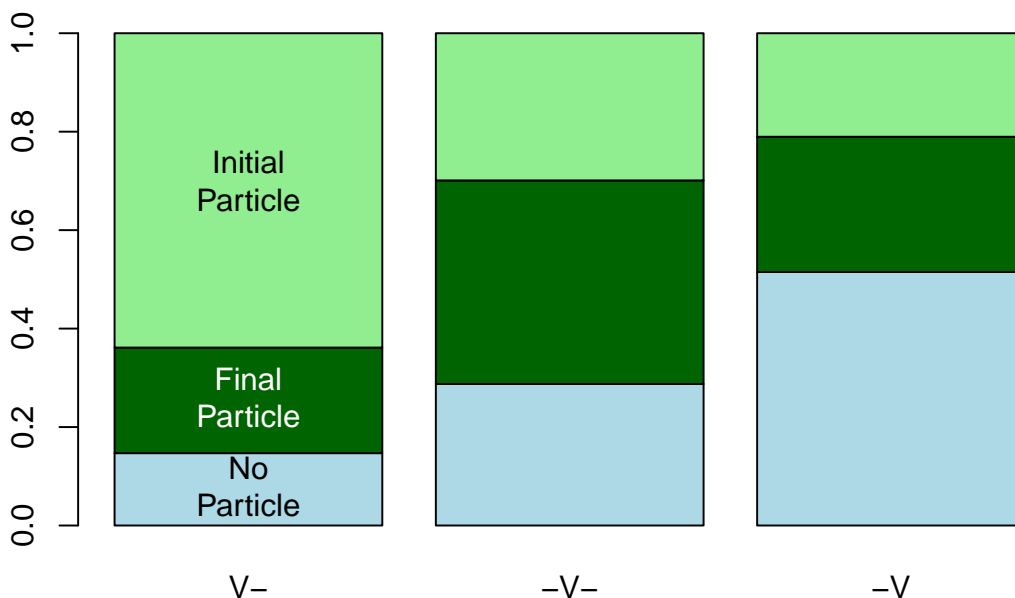
```

Plot the model estimates. They reflect the raw estimates well.

```

barplot(
  t(estX),
  beside=F,
  names.arg=c("V-", "-V-", "-V"),
  col=rev(c("light green", 'dark green', 'light blue')))
text(0.7,0.7,"Initial\nParticle")
text(0.7,0.26,"Final\nParticle",col='white')
text(0.7,0.08,"No\nParticle")

```



Save the results

```
estX2 = t(estX)
rownames(estX2) = c('No Partcile', 'Final Particle', 'Initial Particle')
save(estX2, file='QParticleRealDataEstiamtes.rDat')
```

# Model exploration

## 1 Robustness to turn taking pressure

Figure 1 shows how the distributions of word orders varies with the number of turns in each conversation. Note that the total number of turns remains the same, only the division into conversations differs. For each combination of  $N_{conversation}$  and  $N_{turns}$ , 100 simulations were run ( $N_{agents} = 10, \alpha = 0.1, \beta = 0$ ). Even a small pressure from turn taking creates a difference in word order distributions.

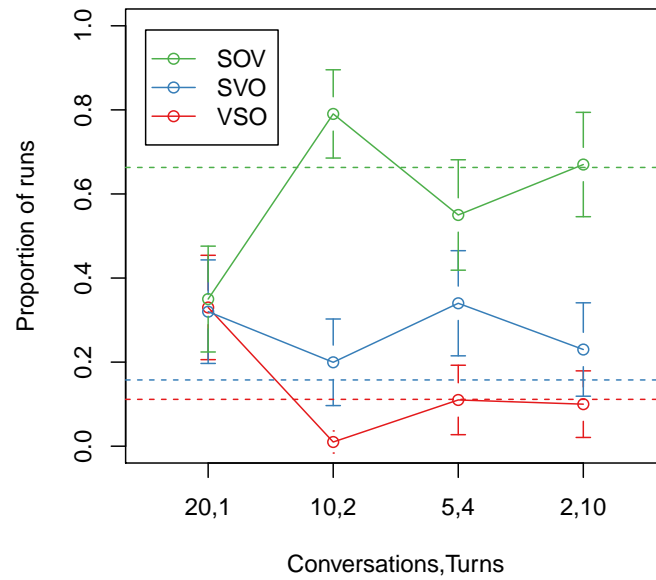


Figure 1: How the distributions of word orders varies with the number of turns in each conversation. The horizontal axis shows number of conversations and number of turns, respectively.

## 2 Effect of population size and noise level

The population size and noise level parameters were explored. Increasing both variables should make the model less susceptible to stochastic effects. The fit of the outcome of the model was compared to the real distributions. This was measured as the sum of the squared error between the proportion of each word order type from the model prediction and from the estimates from real data. Figure 2 shows that the error decreases as the population size and amount of noise increases.

The model implementation is not optimised for fast processing, so only a limited range of population sizes could be tested. The size here is obviously very small compare, but the principles in the main results should hold.

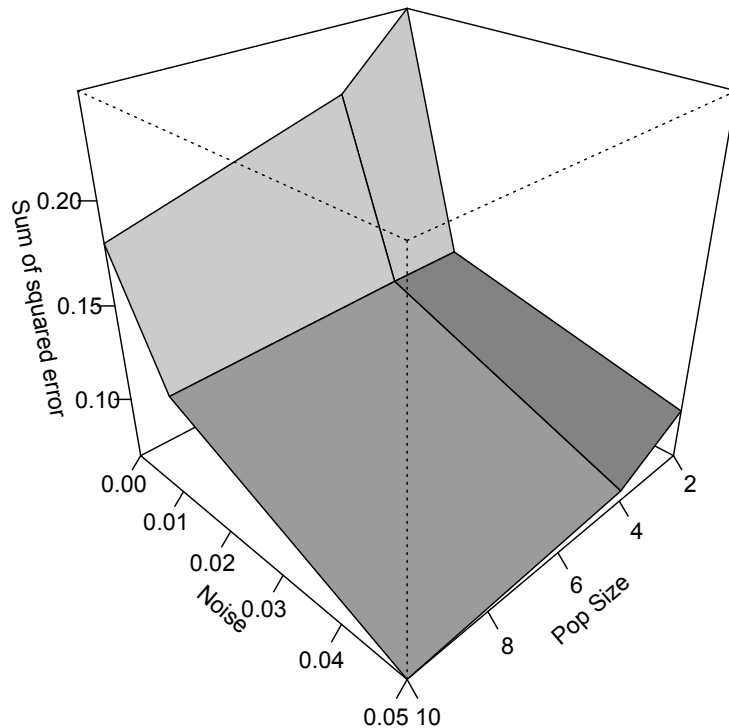


Figure 2: How the fit of the model varies with noise and population size.

### 3 Initial conditions of the model

In the main paper, the agents in the first generation start with a uniform distribution of items in their memory, modelling a situation of no dominant word order. We manipulated this to explore the effects of initial conditions. For each of the basic word orders (SOV, SVO and VSO), we ran simulations where that target word order began with a higher frequency in the first generation’s memory. Simulations were run where the target word order made up 33%, 40%, 60% and 80% of the initial distribution, with the other two types dividing the rest of the distribution evenly. For each setting, 50 simulations were run ( $n_{conversations} = 1, n_{turns} = 20, \beta = 0, \alpha = 0.1$ ).

Figure 3 below shows how the proportion of dominant word orders that arise varies with these settings. With extreme initial conditions, the majority of runs converge on the initially biased order. However, this effect is strongest when the model is initially biased towards SOV, and weakest for an initial bias towards VSO. If SOV makes up 80% of the initial variants, then it goes on to dominate in 98% of runs. For SVO this is 80% and for VSO it is 58%. That is, even when initial conditions are favour VSO order, the pressure for turn taking can reduce the probability of converging on VSO order. The ‘correct’ ranking of word order types is robust to any single word order being up to one third more frequent than the others (second column in figure 3), and is most robust when starting from an SOV-biased condition (top row of figure 3).

This does not happen when the pressures from turn taking are removed ( $n_{conversations} = 20, n_{turns} = 1$ , see figure 4). The only way to recover the ranking of  $SOV > SVO > VSO$  is to assume that initial conditions strongly favoured SOV.

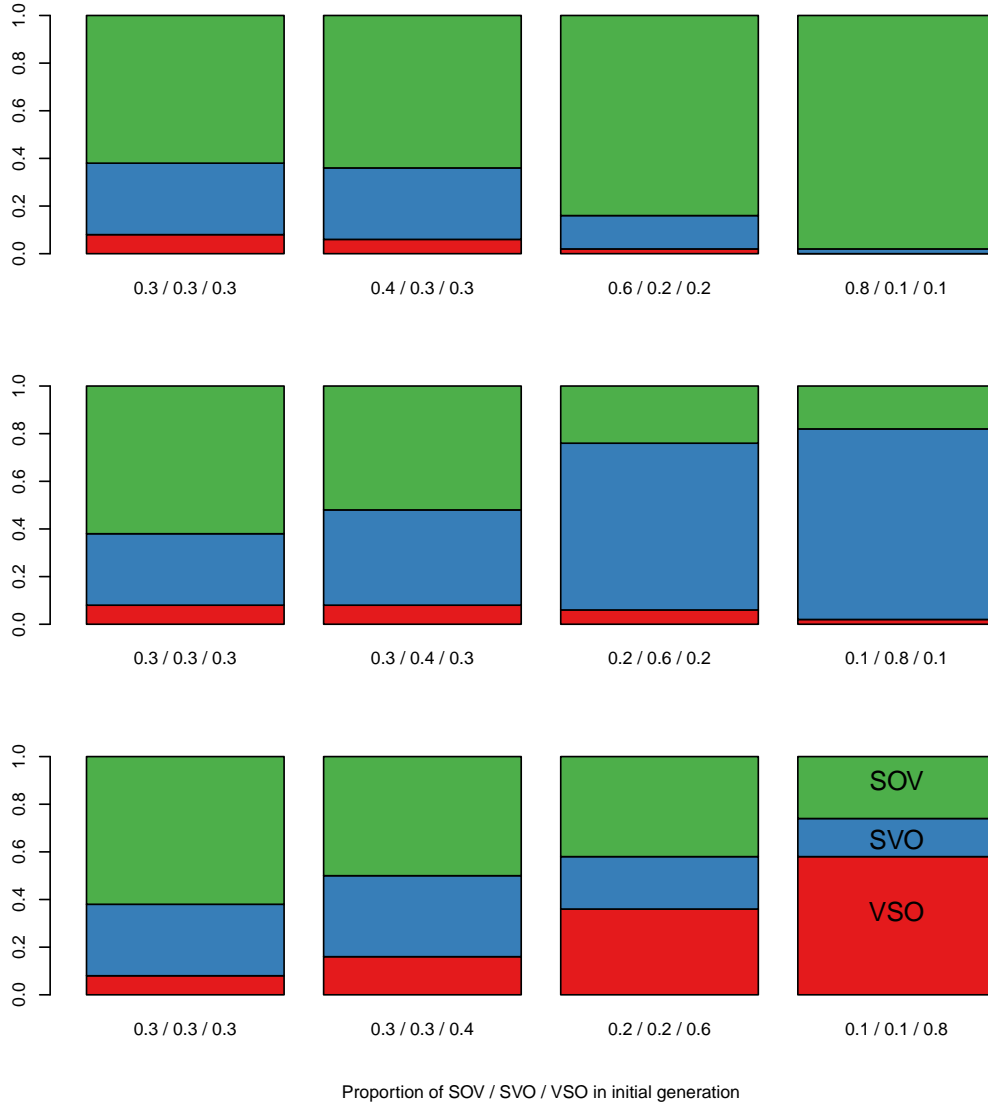


Figure 3: Each stacked bar shows the proportion of runs converging to dominant SOV, SVO and VOS orders for a given condition. Labels on the x axes show the initial proportion of SOV / SVO and VSO orders in the first generation.



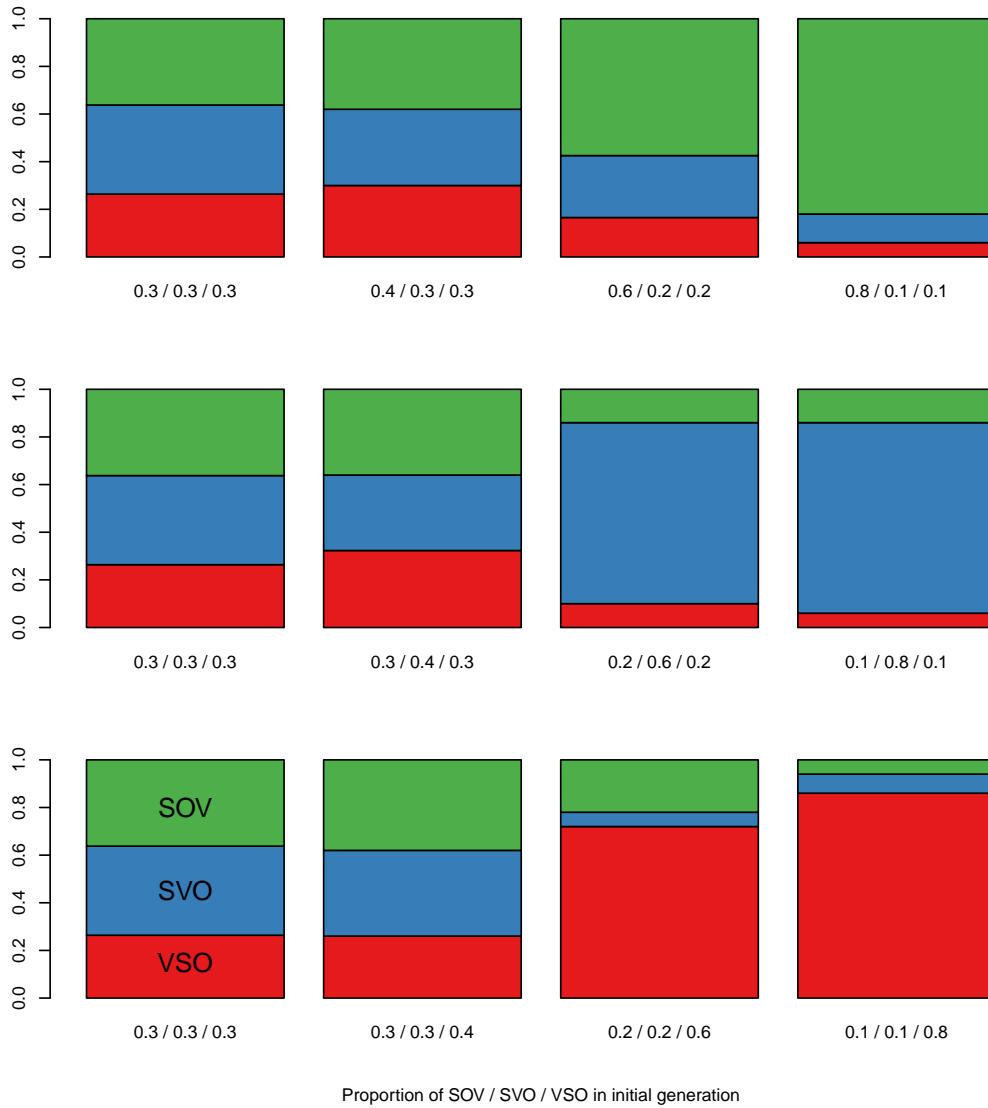


Figure 4: As above, but where the model does not have a pressure for rapid turn taking ( $n_{conversations} = 20, n_{turns} = 1$ ).

## 4 Transitions between states

Gell-Mann and Ruhlen (2011) review historical changes to word order and estimate transitions between languages. They find that word order tends to change from SOV to SVO to VSO, and suggest that languages began as SOV.

We ran many sequences with the standard parameters, observing the dominant order at each generation, where the dominant order was the order accounting for more than 50% of the distribution, or the previous dominant order (this was done to smooth out random fluctuations). We then explored the probability of transitions between each word order.

	VSO	SVO	SOV
VSO	0.9803	0.0080	0.0085
SVO	0.0027	0.9872	0.0068
SOV	0.0014	0.0033	0.9919

Table 1: Transition probabilities between dominant word orders within a run of the model. Columns = from, rows = to, e.g. SVO changes to SOV 0.68% of the time. Each row sums to 1.

The most common pattern is to stay with the existing dominant variant, accounting for about 98% of transitions. SOV is about twice as likely to transition to SVO than VSO, in line with Gell-Mann and Ruhlen’s order. However, SVO is about twice as likely to transition to SOV than to VSO. Also, VSO is about equally likely to transition to either of the other types. In summary, the model transitions do not fit well with those outlined in the linguistic literature on historical change.

However, the mechanism of change discussed in Gell-Mann and Ruhlen is grammaticalisation, rather than drift, and we don’t have any model of interactions between grammatical structures. They also suggest that word order distributions have not reached a stable equilibrium, while in our study we assume that they have. Our model is more focussed on the stage of language evolution that Scott-Phillips and Kirby (2010) call “cultural evolution” - the initial emergence of salient features of language such as dominant word order, rather than “historical change” which describes more recent changes between dominant states. In this case, an alternative view of our model is as an explanation of the starting conditions which feed into the historical changes described in Gell-Mann and Ruhlen. That is, languages are likely to go from no fixed word order to having SOV order. Indeed, Our model does not have any additional mechanisms such as grammaticalisation which would cause transitions between dominant word orders after the initial convergence.

## 4.1 Model with particles

The table below gives the transition probabilities for the model with optional initial or final particles, based on 1000 runs of 300 generations each ( $N_{conversations}=20$ ,  $N_{turns}=1$ ,  $N_{agents} = 10$ ,  $\alpha = 0.1$ ,  $\beta = 0$ ,  $p = 0.5$ ). Numbers are proportions of generations which transition from the type in the vertical row to the type in the horizontal column (e.g. VSO transitions to VSO in 92.2% of generations). Generations are coded based on the most frequent order type. Generations that have no dominant order are coded as "None".

	None	VSO	SVO	SOV	VSOX	SVOX	SOVX	XVSO	XSVO	XSOV
None	0.264	0.042	0.073	0.090	0.052	0.086	0.103	0.075	0.095	0.119
VSO	0.029	0.929	0.004	0.005	0.005	0.004	0.006	0.004	0.006	0.008
SVO	0.017	0.002	0.963	0.003	0.001	0.003	0.003	0.003	0.003	0.003
SOV	0.010	0.000	0.001	0.978	0.001	0.001	0.002	0.001	0.002	0.003
VSOX	0.017	0.002	0.001	0.003	0.957	0.003	0.004	0.004	0.004	0.004
SVOX	0.014	0.001	0.002	0.002	0.001	0.970	0.002	0.002	0.002	0.002
SOVX	0.010	0.001	0.001	0.002	0.001	0.001	0.980	0.001	0.002	0.002
XVSO	0.011	0.001	0.003	0.002	0.001	0.002	0.002	0.975	0.002	0.002
XSVO	0.007	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.984	0.002
XSOV	0.006	0.000	0.001	0.001	0.000	0.001	0.001	0.001	0.001	0.987

## References

- Gell-Mann, M. and Ruhlen, M. (2011). The origin and evolution of word order. *Proceedings of the National Academy of Sciences*, 108(42):17290–17295.
- Scott-Phillips, T. C. and Kirby, S. (2010). Language evolution in the laboratory. *Trends in Cognitive Sciences*, 14(9):411–417.