Combined English and Dutch Analysis

```
library(mgcv)
library(lmtest)
# Code for assessing significance of GAM slopes
source("GAM_derivaties.R")
```

Introduction

Below are the main results for English and Dutch. This analysis tests whether the differences are significant by combining the data and adding a random effect for language.

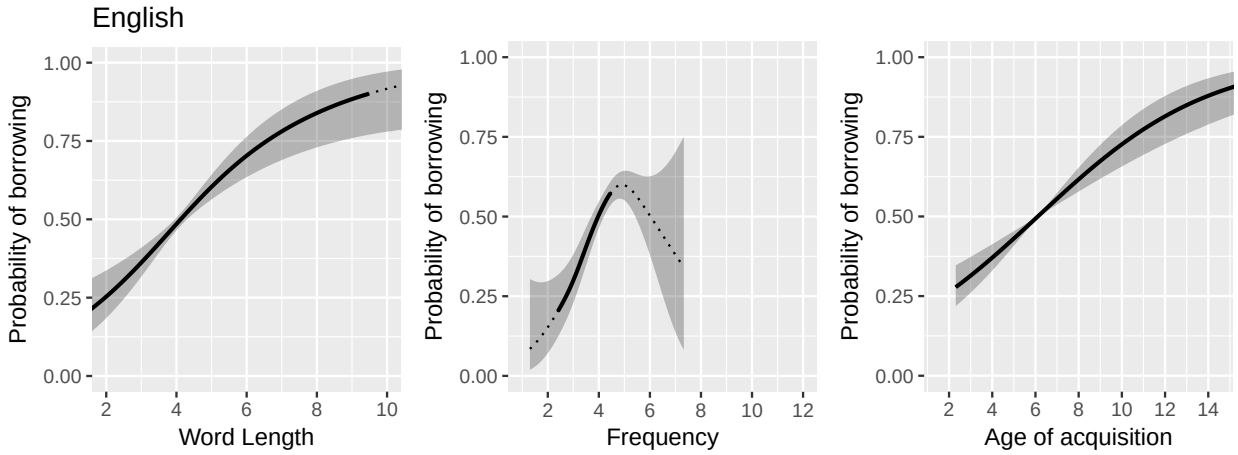


Figure 1: English

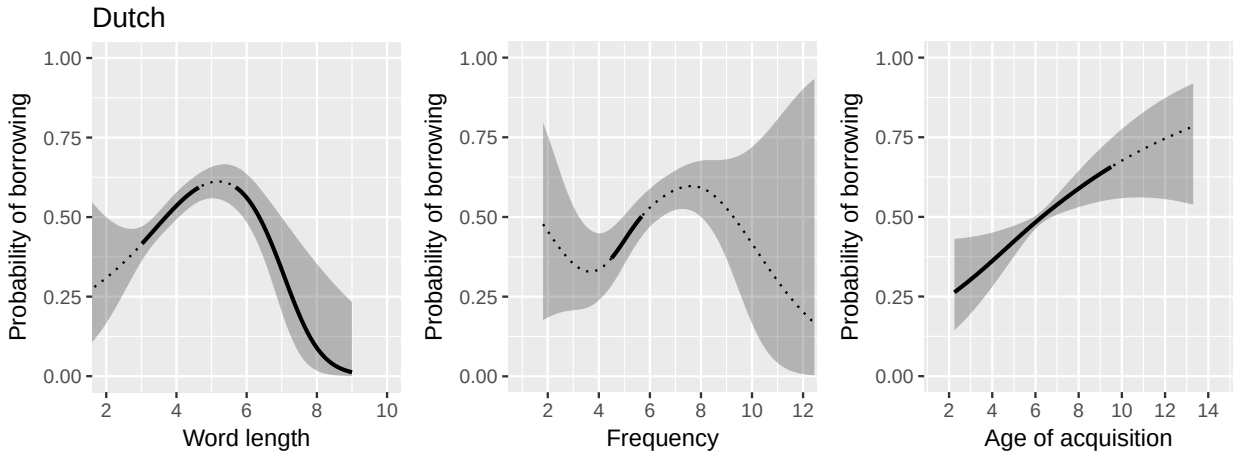


Figure 2: Dutch

Load data

Dutch data:

```
load("../data/loanwords_Dutch.Rdat")
dutch$subtlexzipf =  log10(as.numeric(dutch$subtlex.dominant.pos.frequency))
```

English data:

```
dataloan <- read.csv("../data/loanword12.csv",stringsAsFactors = F)
dataloan$bor15 <- ifelse(dataloan$borrowing==1,1, ifelse(dataloan$borrowing==5,0,NA))
dataloan$bor15.cat <- factor(dataloan$bor15)
dataloan$conc = as.numeric(dataloan$conc)
dataloan$AoA = as.numeric(dataloan$AoA)
dataloan$subtlexzipf = as.numeric(dataloan$subtlexzipf)
```

Make Dutch data frame compatible with English:

```
dataloan$language = "English"
dutch$language = "Dutch"

names(dutch)[names(dutch)=="spelling"] = "word"
names(dutch)[names(dutch)=="length"] = "phonlength"
names(dutch)[names(dutch)=="aoa"] = "AoA"
names(dutch)[names(dutch)=="concreteness"] = "conc"

dutchToEngPos = c(
  N="Noun",
  ADJ="Adjective",
  WW="Verb",
  SPEC="Name",
  VZ="Preposition",
  TW="Number",
  BW="Adverb",
  VG="Cobjuction",
  LID="Interjection",
  TSW="Article",
  VNW="Pronoun")

dutch$cat = dutchToEngPos[as.character(dutch$cat)]
dutch$cat = factor(dutch$cat)
dutch$cat = relevel(dutch$cat,"Noun")
```

Combine data:

```
cols = c("word","bor15.cat","phonlength",
         "AoA", "subtlexzipf",
         "conc","cat","language")

EnglishAndDutch =
  rbind(
    dutch[,cols],
    dataloan[,cols]
  )
EnglishAndDutch$language = factor(EnglishAndDutch$language)
```

```
EnglishAndDutch = EnglishAndDutch[complete.cases(EnglishAndDutch),]
```

Scale and center (note that the data is now centered according to the combined mean):

```
EnglishAndDutch$AoAscale <- scale(EnglishAndDutch$AoA)

EnglishAndDutch$subtlexzipfscale <- scale(EnglishAndDutch$subtlexzipf)

phonlength.center = median(EnglishAndDutch$phonlength)
EnglishAndDutch$phonlengthscale <-
  EnglishAndDutch$phonlength - phonlength.center
phonlength.scale = sd(EnglishAndDutch$phonlengthscale)
EnglishAndDutch$phonlengthscale = EnglishAndDutch$phonlengthscale/phonlength.scale

attr(EnglishAndDutch$phonlengthscale,"scaled:scale") = phonlength.scale
attr(EnglishAndDutch$phonlengthscale,"scaled:center") = phonlength.center

EnglishAndDutch$concscale <- scale(EnglishAndDutch$conc)
conc.scale = attr(EnglishAndDutch$concscale,"scaled:scale")
conc.center = attr(EnglishAndDutch$concscale,"scaled:center")

EnglishAndDutch$cat = relevel(EnglishAndDutch$cat,"Noun")
```

Test of differences between languages

Baseline model with random effect for language:

```
m0 = bam(bor15.cat ~
      s(phonlengthscale) +
      s(AoAscale) +
      s(subtlexzipfscale) +
      s(concscale) +
      s(cat,bs='re')+
      s(cat,phonlengthscale,bs='re')+
      s(cat,AoAscale,bs='re')+
      s(cat,subtlexzipfscale,bs='re')+
      s(cat,concscale,bs='re') +
      s(language,bs='re'),
    data = EnglishAndDutch,
    family='binomial')
```

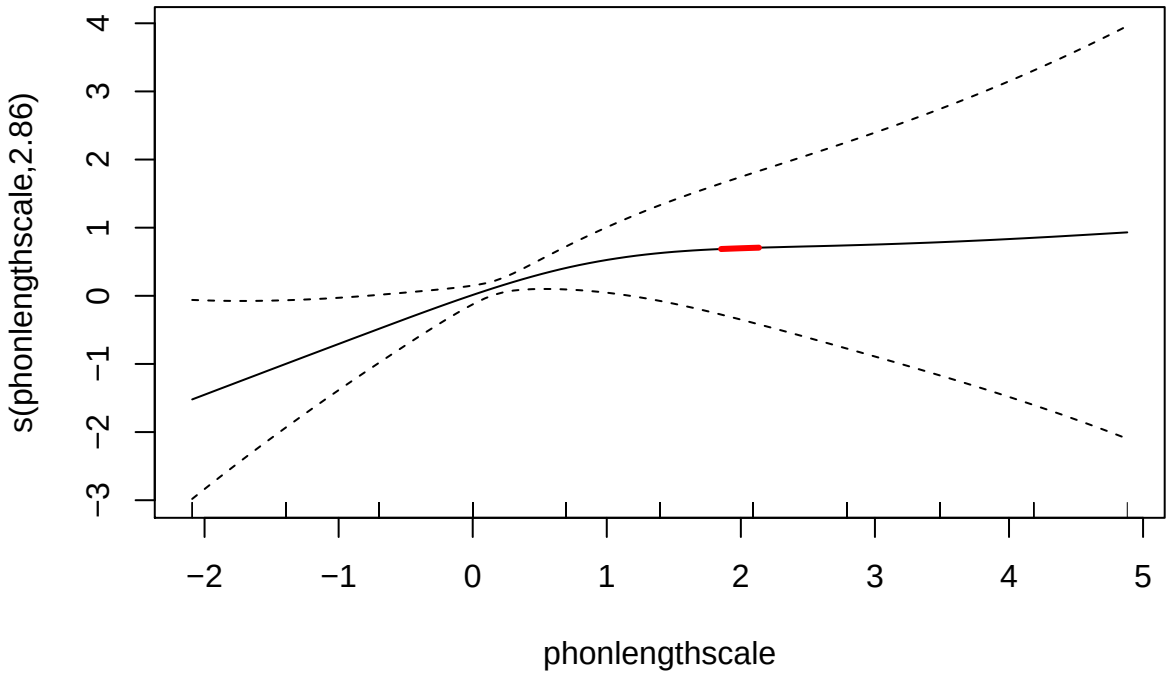
Word Length

```
mLen = update(m0, ~. + s(language, phonlengthscale,bs="re"))
lrtest(m0,mLen)
```

```
## Likelihood ratio test
##
## Model 1: bor15.cat ~ s(phonlengthscale) + s(AoAscale) + s(subtlexzipfscale) +
##     s(concscale) + s(cat, bs = "re") + s(cat, phonlengthscale,
##     bs = "re") + s(cat, AoAscale, bs = "re") + s(cat, subtlexzipfscale,
##     bs = "re") + s(cat, concscale, bs = "re") + s(language, bs = "re")
```

```
## Model 2: bor15.cat ~ s(phonlengthscale) + s(AoAscale) + s(subtlexzipfscale) +
##     s(concscale) + s(cat, bs = "re") + s(cat, phonlengthscale,
##     bs = "re") + s(cat, AoAscale, bs = "re") + s(cat, subtlexzipfscale,
##     bs = "re") + s(cat, concscale, bs = "re") + s(language, bs = "re") +
##     s(language, phonlengthscale, bs = "re")
##      #Df  LogLik     Df Chisq Pr(>Chisq)
## 1 29.566 -1229.5
## 2 30.585 -1220.3 1.0183 18.54  1.663e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
plotGAMSignificantSlopes(mLen,"phonlengthscale","Length")
```

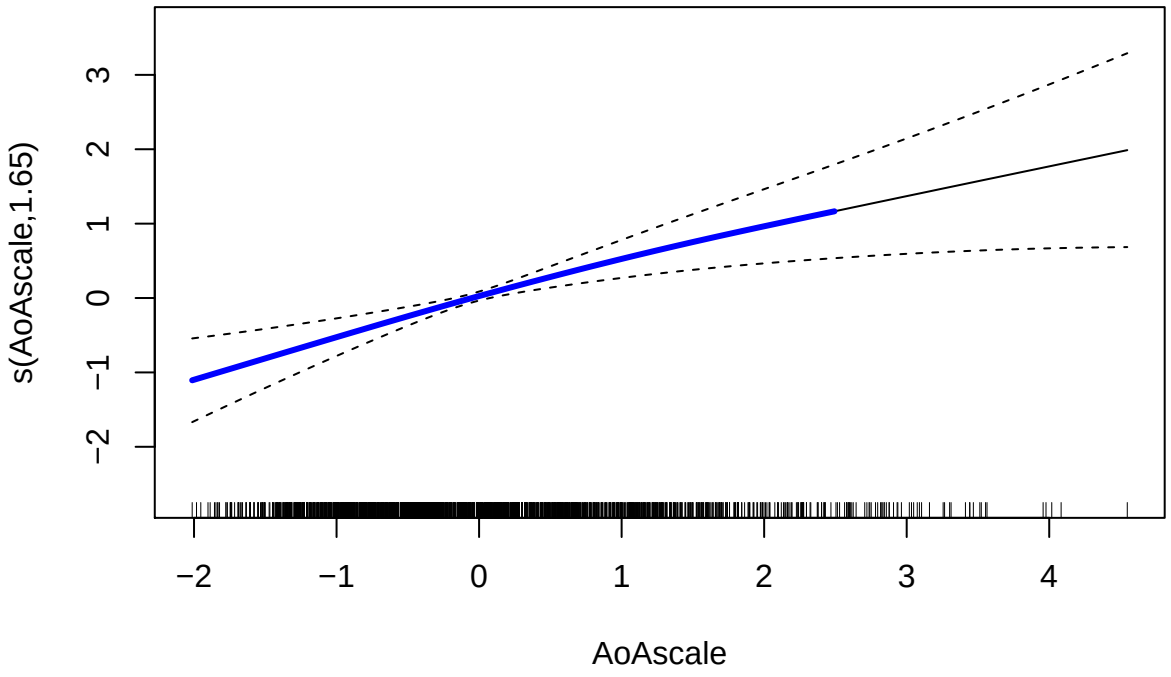


AoA

```
mAoA = update(m0, ~. + s(language, AoAscale,bs="re"))
lrtest(m0,mAoA)
```

```
## Likelihood ratio test
##
## Model 1: bor15.cat ~ s(phonlengthscale) + s(AoAscale) + s(subtlexzipfscale) +
##     s(concscale) + s(cat, bs = "re") + s(cat, phonlengthscale,
##     bs = "re") + s(cat, AoAscale, bs = "re") + s(cat, subtlexzipfscale,
##     bs = "re") + s(cat, concscale, bs = "re") + s(language, bs = "re")
## Model 2: bor15.cat ~ s(phonlengthscale) + s(AoAscale) + s(subtlexzipfscale) +
##     s(concscale) + s(cat, bs = "re") + s(cat, phonlengthscale,
##     bs = "re") + s(cat, AoAscale, bs = "re") + s(cat, subtlexzipfscale,
##     bs = "re") + s(cat, concscale, bs = "re") + s(language, bs = "re") +
##     s(language, AoAscale, bs = "re")
##      #Df  LogLik     Df  Chisq Pr(>Chisq)
## 1 29.566 -1229.5
## 2 31.272 -1229.1 1.7054 0.8659     0.6486
```

```
plotGAMSignificantSlopes(mAoA,"AoAscale","AoA")
```

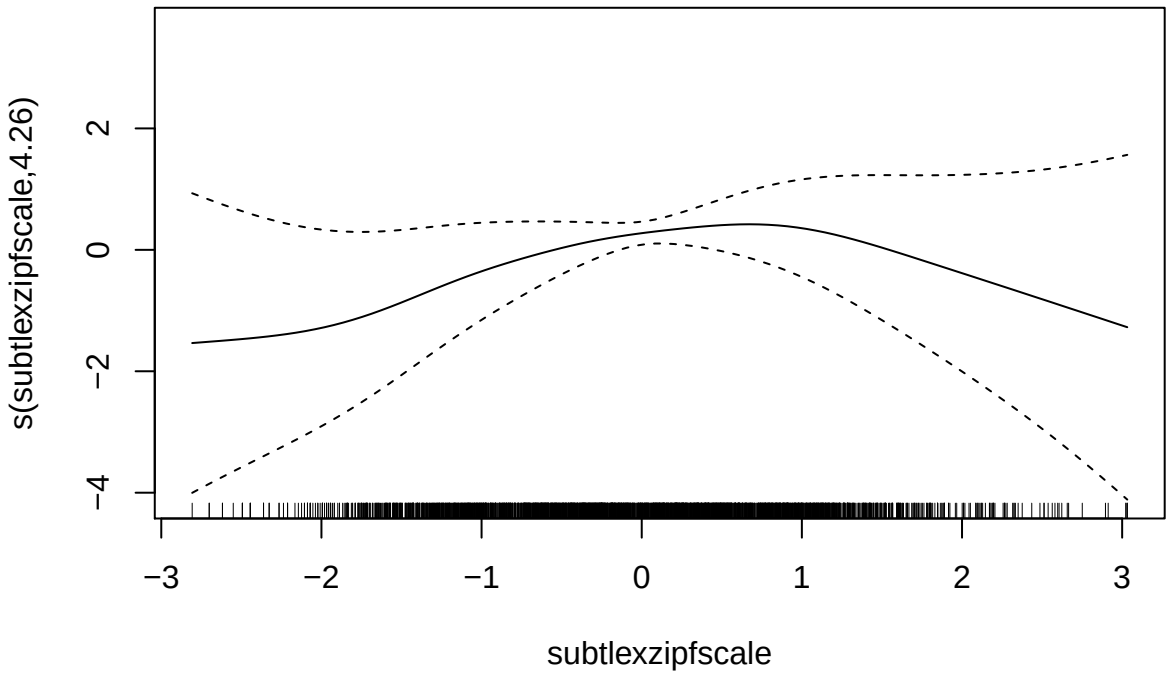


Frequency

```
mFreq = update(m0, ~. + s(language, subtlexzipfscale,bs="re"))
lrtest(m0,mFreq)
```

```
## Likelihood ratio test
##
## Model 1: bor15.cat ~ s(phonlengthscale) + s(AoAscale) + s(subtlexzipfscale) +
##     s(concscale) + s(cat, bs = "re") + s(cat, phonlengthscale,
##     bs = "re") + s(cat, AoAscale, bs = "re") + s(cat, subtlexzipfscale,
##     bs = "re") + s(cat, concscale, bs = "re") + s(language, bs = "re")
## Model 2: bor15.cat ~ s(phonlengthscale) + s(AoAscale) + s(subtlexzipfscale) +
##     s(concscale) + s(cat, bs = "re") + s(cat, phonlengthscale,
##     bs = "re") + s(cat, AoAscale, bs = "re") + s(cat, subtlexzipfscale,
##     bs = "re") + s(cat, concscale, bs = "re") + s(language, bs = "re") +
##     s(language, subtlexzipfscale, bs = "re")
##      #Df  LogLik       Df  Chisq Pr(>Chisq)
## 1 29.566 -1229.5
## 2 29.301 -1224.9 -0.26555 9.2611  < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
plot.gam(mFreq,select=3)
```

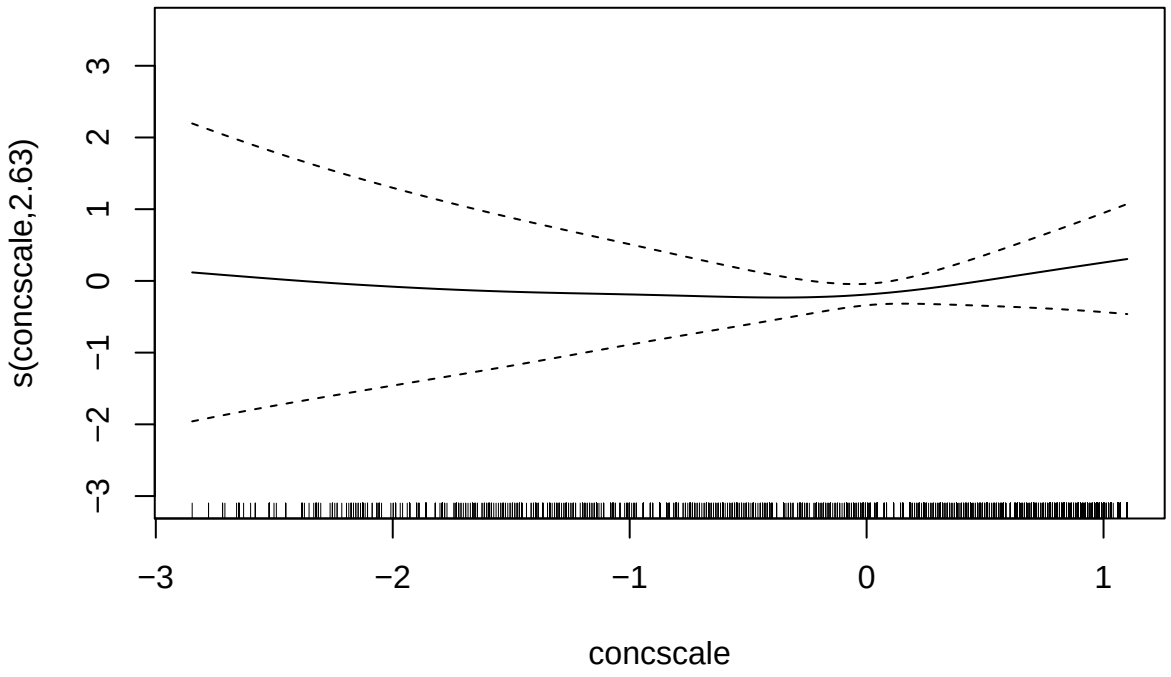


Concreteness

```
mConc = update(m0, ~. + s(language, concscale,bs="re"))
lrtest(m0,mConc)
```

```
## Likelihood ratio test
##
## Model 1: bor15.cat ~ s(phonlengthscale) + s(AoAscale) + s(subtlexzipfscale) +
##     s(concscale) + s(cat, bs = "re") + s(cat, phonlengthscale,
##     bs = "re") + s(cat, AoAscale, bs = "re") + s(cat, subtlexzipfscale,
##     bs = "re") + s(cat, concscale, bs = "re") + s(language, bs = "re")
## Model 2: bor15.cat ~ s(phonlengthscale) + s(AoAscale) + s(subtlexzipfscale) +
##     s(concscale) + s(cat, bs = "re") + s(cat, phonlengthscale,
##     bs = "re") + s(cat, AoAscale, bs = "re") + s(cat, subtlexzipfscale,
##     bs = "re") + s(cat, concscale, bs = "re") + s(language, bs = "re") +
##     s(language, concscale, bs = "re")
##      #Df  LogLik        Df  Chisq Pr(>Chisq)
## 1 29.566 -1229.5
## 2 29.522 -1215.0 -0.043985 29.048  < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
plot.gam(mConc,select=4)
```



Summary

The effect of age of acquisition on borrowing does not differ by language. The effect of word length and frequency on borrowing does differ by language. Adding an interaction between language and concreteness significantly improves the model, suggesting that the relationship between concreteness and borrowing differs by language. However, in the latter case, one should note that there was no significant main effect of concreteness in the main analysis for each language.