A corpus-based analysis of ongoing change in the adjective amplifier systems of Hong Kong, Philippine, and Indian English - Part 2

Martin Schweinberger

2020-08-25

# Introduction

This document shows an analysis of adjective amplification in Irish English based on the Irish component of the International Corpus of English (ICE).

## Preparation

In a first step, the session is prepared by clearing the work space, setting options, activating packages and functions, as well as loading relevant functions.

# clean current workspace  
rm(list=ls(all=T))   
# load packages  
library(tidyverse)

## Error in library(tidyverse): there is no package called 'tidyverse'

library(dplyr)

## Error in library(dplyr): there is no package called 'dplyr'

library(tidyr)

## Error in library(tidyr): there is no package called 'tidyr'

library(ggplot2)

## Error in library(ggplot2): there is no package called 'ggplot2'

# set options  
options(stringsAsFactors = F)   
options(scipen = 999)   
options(max.print=10000)  
# define image directory  
imageDirectory <- "images"

## Load data

ice <- read.delim("D:\\Uni\\Projekte\\02-Intensification\\AmpHKPIE\\data\\editdata//iceamp\_04\_clean.txt", sep = "\t", header = T)  
# inspect  
head(ice)

## Id Corpus File Speaker  
## 1 20 ICE-HK S1A-001 A  
## 2 34 ICE-HK S1A-001 B  
## 3 35 ICE-HK S1A-001 A  
## 4 36 ICE-HK S1A-001 A  
## 5 49 ICE-HK S1A-001 A  
## 6 58 ICE-HK S1A-001 A  
## SpeechUnit  
## 1 Uhm <,> the youngest is supposed to do the least <,> housework but actually <O> $Z-laughs </O> uhm <,> but actually the fact is not <{1> <[1> <,> </[1> quite the same because uhm my two brothers <O> laughs </O> always <,> always said that they are <,> I am <?> already </?> so busy   
## 2 Very long distance <$Z> <X>   
## 3 I 'm very lazy actually <$Z> <X>   
## 4 <[1> I 'm very lazy </[1> </{1>   
## 5 <[> It 's </[> </{> it 's the Chinese medicine for <,> uhm <.> s </.> uhm <,> for the <,> for the illness of <,> stomach <{1> <[1> <,> </[1> but <,> but it is very traditional because uhm <,> they they are arranged in a <,> pocket <{2> <[2> and </[2> when there <,> is something like <,> uhm <,> dusts <$Z> <X>   
## 6 A very interesting story   
## SpeechUnitCount WordCount Date Gender Age Nationality L1  
## 1 250 2213 1996 female 17-25 Chinese Cantonese  
## 2 49 244 1996 female 17-25 Chinese Cantonese  
## 3 250 2213 1996 female 17-25 Chinese Cantonese  
## 4 250 2213 1996 female 17-25 Chinese Cantonese  
## 5 250 2213 1996 female 17-25 Chinese Cantonese  
## 6 250 2213 1996 female 17-25 Chinese Cantonese  
## EducationLevel OtherLanguages Birthplace Ethnicity PreContext Token  
## 1 NoCollege English Hong Kong Chinese so/RB busy/JJ  
## 2 NoCollege English Hong Kong Chinese Very/RB long/JJ  
## 3 NoCollege English Hong Kong Chinese very/RB lazy/JJ  
## 4 NoCollege English Hong Kong Chinese very/RB lazy/JJ  
## 5 NoCollege English Hong Kong Chinese very/RB traditional/JJ  
## 6 NoCollege English Hong Kong Chinese very/RB interesting/JJ  
## PostContext Adjective Variant Function  
## 1 busy so Predicative  
## 2 distance/NN long very Attributive  
## 3 actually/RB lazy very Predicative  
## 4 lazy very Predicative  
## 5 because/IN uhm/NN they/PR traditional very Predicative  
## 6 story/NN interesting very Attributive  
## PreContextLong Amplified Frequency Correction Priming  
## 1 RP am/VBP already/RB so/RB 1 0.50 correct NotPrimed  
## 2 Very/RB 1 1.07 correct NotPrimed  
## 3 I/PRP ' m/VBP very/RB 1 0.18 correct Primed  
## 4 I/PRP ' m/VBP very/RB 1 0.18 correct Primed  
## 5 t/CC it/PRP is/VBZ very/RB 1 0.13 correct Primed  
## 6 A/DT very/RB 1 0.74 correct NotPrimed  
## Gradability SemanticCategory Emotionality  
## 1 -3.6635616 HumanPropensity NonEmotional  
## 2 -0.8156624 Dimension NonEmotional  
## 3 -0.3132649 HumanPropensity NegativeEmotional  
## 4 -1.6660953 HumanPropensity NegativeEmotional  
## 5 1.0000000 Other PositiveEmotional  
## 6 1.0000000 Value PositiveEmotional

## Clean data

Factorize predictors and scale frequency

ice <- ice %>%  
 select(-Id, -Correction, -SpeechUnit, -SpeechUnitCount, -PreContext, -Token, -L1,   
 -PostContext, -PreContextLong, -Amplified, -Nationality, -OtherLanguages, -Birthplace) %>%  
 dplyr::mutate(Emotionality = ifelse(Emotionality == "PositiveEmotional", "Positive",  
 ifelse(Emotionality == "NonEmotional", "Neutral", "Negative")),  
 Emotionality = factor(Emotionality, levels = c("Negative", "Neutral", "Positive")),  
 Frequency = as.vector(scale(Frequency)),  
 Gender = factor(Gender),  
 EducationLevel = factor(EducationLevel),  
 Priming = factor(Priming),  
 Emotionality = factor(Emotionality),  
 Function = factor(Function),  
 Ethnicity = factor(Ethnicity),  
 SemanticCategory = factor(SemanticCategory),  
 Speaker = paste0(File, "$", Speaker)) %>%  
 select(-File, -WordCount)

## Error in ice %>% select(-Id, -Correction, -SpeechUnit, -SpeechUnitCount, : konnte Funktion "%>%" nicht finden

str(ice)

## 'data.frame': 3126 obs. of 30 variables:  
## $ Id : int 20 34 35 36 49 58 67 75 89 100 ...  
## $ Corpus : chr "ICE-HK" "ICE-HK" "ICE-HK" "ICE-HK" ...  
## $ File : chr "S1A-001" "S1A-001" "S1A-001" "S1A-001" ...  
## $ Speaker : chr "A" "B" "A" "A" ...  
## $ SpeechUnit : chr "Uhm <,> the youngest is supposed to do the least <,> housework but actually <O> $Z-laughs </O> uhm <,> but actu"| \_\_truncated\_\_ "Very long distance <$Z> <X> " "I 'm very lazy actually <$Z> <X> " "<[1> I 'm very lazy </[1> </{1> " ...  
## $ SpeechUnitCount : int 250 49 250 250 250 250 250 250 250 250 ...  
## $ WordCount : int 2213 244 2213 2213 2213 2213 2213 2213 2213 2213 ...  
## $ Date : int 1996 1996 1996 1996 1996 1996 1996 1996 1996 1996 ...  
## $ Gender : chr "female" "female" "female" "female" ...  
## $ Age : chr "17-25" "17-25" "17-25" "17-25" ...  
## $ Nationality : chr "Chinese" "Chinese" "Chinese" "Chinese" ...  
## $ L1 : chr "Cantonese" "Cantonese" "Cantonese" "Cantonese" ...  
## $ EducationLevel : chr "NoCollege" "NoCollege" "NoCollege" "NoCollege" ...  
## $ OtherLanguages : chr "English" "English" "English" "English" ...  
## $ Birthplace : chr "Hong Kong" "Hong Kong" "Hong Kong" "Hong Kong" ...  
## $ Ethnicity : chr "Chinese" "Chinese" "Chinese" "Chinese" ...  
## $ PreContext : chr "so/RB" "Very/RB" "very/RB" "very/RB" ...  
## $ Token : chr "busy/JJ" "long/JJ" "lazy/JJ" "lazy/JJ" ...  
## $ PostContext : chr "" "distance/NN" "actually/RB" "" ...  
## $ Adjective : chr "busy" "long" "lazy" "lazy" ...  
## $ Variant : chr "so" "very" "very" "very" ...  
## $ Function : chr "Predicative" "Attributive" "Predicative" "Predicative" ...  
## $ PreContextLong : chr "RP am/VBP already/RB so/RB" "Very/RB" "I/PRP ' m/VBP very/RB" "I/PRP ' m/VBP very/RB" ...  
## $ Amplified : int 1 1 1 1 1 1 1 1 1 1 ...  
## $ Frequency : num 0.5 1.07 0.18 0.18 0.13 0.74 0.78 0.07 0.01 0.8 ...  
## $ Correction : chr "correct" "correct" "correct" "correct" ...  
## $ Priming : chr "NotPrimed" "NotPrimed" "Primed" "Primed" ...  
## $ Gradability : num -3.664 -0.816 -0.313 -1.666 1 ...  
## $ SemanticCategory: chr "HumanPropensity" "Dimension" "HumanPropensity" "HumanPropensity" ...  
## $ Emotionality : chr "NonEmotional" "NonEmotional" "NegativeEmotional" "NegativeEmotional" ...

Determine adjective frequency

fadj <- names(table(ice$Adjective))[which(table(ice$Adjective) > 15)]  
fadj

## [1] "bad" "beautiful" "big" "boring" "busy"   
## [6] "cheap" "cold" "dangerous" "different" "difficult"   
## [11] "easy" "expensive" "famous" "funny" "good"   
## [16] "great" "happy" "hard" "high" "hot"   
## [21] "important" "interesting" "long" "low" "lucky"   
## [26] "nice" "old" "sad" "serious" "simple"   
## [31] "small" "strange" "strong" "sweet" "weird"   
## [36] "young"

Determine variant frequency

famp <- names(table(ice$Variant))[which(table(ice$Variant) > 50)]  
famp

## [1] "really" "so" "very"

Collapse infrequent adjectives and amplifiers

ice <- ice %>%  
 mutate(Adjective = ifelse(Adjective %in% fadj, Adjective, "other"),  
 Variant = ifelse(Variant %in% famp, Variant, "other"),  
 Adjective = factor(Adjective),  
 Variant = factor(Variant)) %>%  
 group\_by(Corpus) %>%  
 mutate(Frequency = scale(Frequency)[,1])

## Error in ice %>% mutate(Adjective = ifelse(Adjective %in% fadj, Adjective, : konnte Funktion "%>%" nicht finden

str(ice)

## 'data.frame': 3126 obs. of 30 variables:  
## $ Id : int 20 34 35 36 49 58 67 75 89 100 ...  
## $ Corpus : chr "ICE-HK" "ICE-HK" "ICE-HK" "ICE-HK" ...  
## $ File : chr "S1A-001" "S1A-001" "S1A-001" "S1A-001" ...  
## $ Speaker : chr "A" "B" "A" "A" ...  
## $ SpeechUnit : chr "Uhm <,> the youngest is supposed to do the least <,> housework but actually <O> $Z-laughs </O> uhm <,> but actu"| \_\_truncated\_\_ "Very long distance <$Z> <X> " "I 'm very lazy actually <$Z> <X> " "<[1> I 'm very lazy </[1> </{1> " ...  
## $ SpeechUnitCount : int 250 49 250 250 250 250 250 250 250 250 ...  
## $ WordCount : int 2213 244 2213 2213 2213 2213 2213 2213 2213 2213 ...  
## $ Date : int 1996 1996 1996 1996 1996 1996 1996 1996 1996 1996 ...  
## $ Gender : chr "female" "female" "female" "female" ...  
## $ Age : chr "17-25" "17-25" "17-25" "17-25" ...  
## $ Nationality : chr "Chinese" "Chinese" "Chinese" "Chinese" ...  
## $ L1 : chr "Cantonese" "Cantonese" "Cantonese" "Cantonese" ...  
## $ EducationLevel : chr "NoCollege" "NoCollege" "NoCollege" "NoCollege" ...  
## $ OtherLanguages : chr "English" "English" "English" "English" ...  
## $ Birthplace : chr "Hong Kong" "Hong Kong" "Hong Kong" "Hong Kong" ...  
## $ Ethnicity : chr "Chinese" "Chinese" "Chinese" "Chinese" ...  
## $ PreContext : chr "so/RB" "Very/RB" "very/RB" "very/RB" ...  
## $ Token : chr "busy/JJ" "long/JJ" "lazy/JJ" "lazy/JJ" ...  
## $ PostContext : chr "" "distance/NN" "actually/RB" "" ...  
## $ Adjective : chr "busy" "long" "lazy" "lazy" ...  
## $ Variant : chr "so" "very" "very" "very" ...  
## $ Function : chr "Predicative" "Attributive" "Predicative" "Predicative" ...  
## $ PreContextLong : chr "RP am/VBP already/RB so/RB" "Very/RB" "I/PRP ' m/VBP very/RB" "I/PRP ' m/VBP very/RB" ...  
## $ Amplified : int 1 1 1 1 1 1 1 1 1 1 ...  
## $ Frequency : num 0.5 1.07 0.18 0.18 0.13 0.74 0.78 0.07 0.01 0.8 ...  
## $ Correction : chr "correct" "correct" "correct" "correct" ...  
## $ Priming : chr "NotPrimed" "NotPrimed" "Primed" "Primed" ...  
## $ Gradability : num -3.664 -0.816 -0.313 -1.666 1 ...  
## $ SemanticCategory: chr "HumanPropensity" "Dimension" "HumanPropensity" "HumanPropensity" ...  
## $ Emotionality : chr "NonEmotional" "NonEmotional" "NegativeEmotional" "NegativeEmotional" ...

## Data inspection

Visualize Variant by Corpus and Age

ice %>%  
 mutate(Age = ifelse(Age == "16-25"|Age == "17-25"|Age == "18-25", 3,   
 ifelse(Age == "26-40"|Age == "26-41", 2,   
 ifelse(Age == "41+"|Age == "42+", 1, Age))),  
 Corpus = ifelse(Corpus == "ICE-HK", "ICE Hong Kong",  
 ifelse(Corpus == "ICE-IND", "ICE India", "ICE Philippines"))) %>%  
 group\_by(Corpus, Age, Variant) %>%  
 summarise(Frequency = n()) %>%  
 group\_by(Corpus, Age) %>%  
 mutate(AllSlots = sum(Frequency),  
 Percent = round(Frequency/AllSlots\*100, 1)) %>%  
 select(-Frequency, -AllSlots) %>%  
 ggplot(aes(x = Age, y = Percent,   
 group = Variant, color = Variant, linetype = Variant)) +  
 facet\_grid(cols = vars(Corpus)) +  
 geom\_line() +   
 coord\_cartesian(ylim = c(0, 100)) +  
 theme\_set(theme\_bw(base\_size = 12)) +  
 theme(legend.position="top",   
 panel.grid.major = element\_blank(),   
 panel.grid.minor = element\_blank(),   
 axis.text.x = element\_text(size=10)) +  
 scale\_color\_manual(breaks = c("other", "really", "so", "very"),  
 values = c("gray80", "gray60", "gray40", "gray20")) +  
 scale\_linetype\_manual(breaks = c("other", "really", "so", "very"),  
 values = c(4, 3, 2, 1)) +   
 scale\_x\_discrete(breaks = 1:3,   
 labels = c("41+", "26-41", "16-25")) +  
 ggsave(file =   
 "D:\\Uni\\Projekte\\02-Intensification\\AmpHKPIE\\images/VariantAgeCorpus.png",  
 height = 4, width = 6, dpi = 320)

## Error in ice %>% mutate(Age = ifelse(Age == "16-25" | Age == "17-25" | : konnte Funktion "%>%" nicht finden

Visualize Variant by Corpus and Age and Function

ice %>%  
 mutate(Age = ifelse(Age == "16-25"|Age == "17-25"|Age == "18-25", 3,   
 ifelse(Age == "26-40"|Age == "26-41", 2,   
 ifelse(Age == "41+"|Age == "42+", 1, Age))),  
 Corpus = ifelse(Corpus == "ICE-HK", "ICE Hong Kong",  
 ifelse(Corpus == "ICE-IND", "ICE India", "ICE Philippines"))) %>%  
 group\_by(Corpus, Age, Variant, Function) %>%  
 summarise(Frequency = n()) %>%  
 group\_by(Corpus, Age, Function) %>%  
 mutate(AllSlots = sum(Frequency),  
 Percent = round(Frequency/AllSlots\*100, 1)) %>%  
 select(-Frequency, -AllSlots) %>%  
 ggplot(aes(x = Age, y = Percent,   
 group = Variant, color = Variant, linetype = Variant)) +  
 facet\_grid(cols = vars(Corpus), rows=vars(Function)) +  
 geom\_line() +   
 coord\_cartesian(ylim = c(0, 100)) +  
 theme\_set(theme\_bw(base\_size = 12)) +  
 theme(legend.position="top",   
 panel.grid.major = element\_blank(),   
 panel.grid.minor = element\_blank(),   
 axis.text.x = element\_text(size=10)) +  
 scale\_color\_manual(breaks = c("other", "really", "so", "very"),  
 values = c("gray80", "gray60", "gray40", "gray20")) +  
 scale\_linetype\_manual(breaks = c("other", "really", "so", "very"),  
 values = c(4, 3, 2, 1)) +   
 scale\_x\_discrete(breaks = 1:3,   
 labels = c("41+", "26-41", "16-25")) +  
 ggsave(file =   
 "D:\\Uni\\Projekte\\02-Intensification\\AmpHKPIE\\images/VariantAgeCorpusFunction.png",  
 height = 4, width = 6, dpi = 320)

## Error in ice %>% mutate(Age = ifelse(Age == "16-25" | Age == "17-25" | : konnte Funktion "%>%" nicht finden

# Analysis of ICE-PHI

phi <- ice %>%  
 ungroup() %>%  
 filter(Corpus == "ICE-PHI") %>%  
 mutate(Age = factor(Age),  
 Date = ifelse(Date == "1991"|Date == "1992", "Before1993", "After1998"),  
 Date = factor(Date, levels = c("Before1993", "After1998")),  
 Corpus = factor(Corpus),  
 Speaker = factor(Speaker)) %>%  
 droplevels()

## Error in ice %>% ungroup() %>% filter(Corpus == "ICE-PHI") %>% mutate(Age = factor(Age), : konnte Funktion "%>%" nicht finden

# inspect  
head(phi)

## Error in head(phi): Objekt 'phi' nicht gefunden

Find variables that cannot be tested

l <- sapply(phi, function(x) is.factor(x))

## Error in lapply(X = X, FUN = FUN, ...): Objekt 'phi' nicht gefunden

m <- phi[, l]

## Error in eval(expr, envir, enclos): Objekt 'phi' nicht gefunden

ifelse(n <- sapply(m, function(x) length(levels(x))) == 1, "DROP", "NODROP")

## Error in lapply(X = X, FUN = FUN, ...): Objekt 'm' nicht gefunden

Remove variables and NA

nrow(phi)

## Error in nrow(phi): Objekt 'phi' nicht gefunden

phi\_rfd <- phi %>%  
 select(-Ethnicity, -Corpus, -EducationLevel, -Speaker) %>%  
 na.omit()

## Error in phi %>% select(-Ethnicity, -Corpus, -EducationLevel, -Speaker) %>% : konnte Funktion "%>%" nicht finden

nrow(phi\_rfd)

## Error in nrow(phi\_rfd): Objekt 'phi\_rfd' nicht gefunden

## Random forest

Identify variables to include in regression modeling.

library(randomForest)

## Error in library(randomForest): there is no package called 'randomForest'

set.seed(2020082401)  
rf\_phi <- randomForest(Variant ~., data = phi\_rfd, ntree=10000, proximity=TRUE, importance=TRUE)

## Error in randomForest(Variant ~ ., data = phi\_rfd, ntree = 10000, proximity = TRUE, : konnte Funktion "randomForest" nicht finden

# inspect rf results  
rf\_phi

## Error in eval(expr, envir, enclos): Objekt 'rf\_phi' nicht gefunden

Plot out-of-bag error frequency

# plot new precision/error rate  
oob.error.data <- data.frame(  
 Trees=rep(1:nrow(rf\_phi$err.rate), times=5),  
 Type=rep(c("OOB", "other", "really", "so", "very"), each=nrow(rf\_phi$err.rate)),  
 Error=c(rf\_phi$err.rate[,"OOB"],  
 rf\_phi$err.rate[,"other"],  
 rf\_phi$err.rate[,"really"],  
 rf\_phi$err.rate[,"so"],  
 rf\_phi$err.rate[,"very"]))

## Error in nrow(rf\_phi$err.rate): Objekt 'rf\_phi' nicht gefunden

ggplot(data=oob.error.data, aes(x=Trees, y=Error)) +  
 geom\_line(aes(color=Type))

## Error in ggplot(data = oob.error.data, aes(x = Trees, y = Error)): konnte Funktion "ggplot" nicht finden

Check accuracy and improvement compared to base-line model

# determine accuracy by prediction  
library(caret)

## Error in library(caret): there is no package called 'caret'

# prediction  
pnsrf <- predict(rf\_phi, phi\_rfd)

## Error in predict(rf\_phi, phi\_rfd): Objekt 'rf\_phi' nicht gefunden

confusionMatrix(pnsrf, phi\_rfd$Variant)

## Error in confusionMatrix(pnsrf, phi\_rfd$Variant): konnte Funktion "confusionMatrix" nicht finden

# calculate increase in prediction accuracy compared to base-line model  
cm <- confusionMatrix(pnsrf, phi\_rfd$Variant)

## Error in confusionMatrix(pnsrf, phi\_rfd$Variant): konnte Funktion "confusionMatrix" nicht finden

round(cm[3]$overall[[1]]/cm[3]$overall[[5]]\*100, 2) - 100

## Error in cm[3]: Objekt des Typs 'closure' ist nicht indizierbar

Visualize variable importance

varImpPlot(rf\_phi, main = "", pch = 20, cex = 1.5)

## Error in varImpPlot(rf\_phi, main = "", pch = 20, cex = 1.5): konnte Funktion "varImpPlot" nicht finden

png("D:\\Uni\\Projekte\\02-Intensification\\AmpHKPIE\\images//VarImp\_IcePhi.png", width = 750, height = 500)  
varImpPlot(rf\_phi, main = "", pch = 20, cex = 1.25)

## Error in varImpPlot(rf\_phi, main = "", pch = 20, cex = 1.25): konnte Funktion "varImpPlot" nicht finden

dev.off()

## png   
## 2

### Extract RF-based predictions

pphi\_rf <- predict(rf\_phi, phi\_rfd, type = "prob")

## Error in predict(rf\_phi, phi\_rfd, type = "prob"): Objekt 'rf\_phi' nicht gefunden

phi\_pred <- cbind(phi\_rfd, pphi\_rf)

## Error in cbind(phi\_rfd, pphi\_rf): Objekt 'phi\_rfd' nicht gefunden

pphi <- phi\_pred %>%  
 gather(PredictedVariant, Probability, other:very)

## Error in phi\_pred %>% gather(PredictedVariant, Probability, other:very): konnte Funktion "%>%" nicht finden

# inspect  
head(pphi)

## Error in head(pphi): Objekt 'pphi' nicht gefunden

Visualize predictions

pphi %>%  
 select(Age, Probability, Function, Gender, PredictedVariant) %>%  
 mutate(Age = ifelse(Age == "16-25", 3,   
 ifelse(Age == "26-40", 2,   
 ifelse(Age == "41+", 1, Age)))) %>%  
 group\_by(Age, Function, Gender, PredictedVariant) %>%  
 summarise(Probability = mean(Probability)) %>%  
 ggplot(aes(x = Age, y = Probability,   
 group = PredictedVariant, color = PredictedVariant, linetype = PredictedVariant)) +  
 geom\_line() +   
 facet\_grid(cols = vars(Gender), rows = vars(Function)) +  
 theme\_set(theme\_bw(base\_size = 12)) +  
 theme(legend.position="top",   
 panel.grid.major = element\_blank(),   
 panel.grid.minor = element\_blank(),   
 axis.text.x = element\_text(size=10)) +  
 scale\_color\_manual(breaks = c("other", "really", "so", "very"),  
 values = c("gray80", "gray60", "gray40", "gray20")) +  
 scale\_linetype\_manual(breaks = c("other", "really", "so", "very"),  
 values = c(4, 3, 2, 1)) +   
 scale\_x\_continuous(breaks = 1:3,   
 labels = c("41+", "26-40", "16-25")) +  
 ggsave(file =   
 "D:\\Uni\\Projekte\\02-Intensification\\AmpHKPIE\\images/VariantAgeFunction\_Prob\_Phi.png",  
 height = 5, width = 6, dpi = 320)

## Error in pphi %>% select(Age, Probability, Function, Gender, PredictedVariant) %>% : konnte Funktion "%>%" nicht finden

## Regression analysis

Prepare data

pphi <- pphi %>%  
 select(-Variant) %>%  
 rename(Variant = PredictedVariant)

## Error in pphi %>% select(-Variant) %>% rename(Variant = PredictedVariant): konnte Funktion "%>%" nicht finden

library(rms)

## Error in library(rms): there is no package called 'rms'

library(lme4)

## Error in library(lme4): there is no package called 'lme4'

library(car)

## Error in library(car): there is no package called 'car'

# generate models  
m0.glm <- glm(Probability ~ 1, family = gaussian, data = pphi)

## Error in is.data.frame(data): Objekt 'pphi' nicht gefunden

m0.glmer = glmer(Probability ~ 1 + (1|Adjective), data = pphi, family = "gaussian")

## Error in glmer(Probability ~ 1 + (1 | Adjective), data = pphi, family = "gaussian"): konnte Funktion "glmer" nicht finden

# extract aics  
aic.m0.glmer <- AIC(logLik(m0.glmer))

## Error in logLik(m0.glmer): Objekt 'm0.glmer' nicht gefunden

aic.glm <- AIC(logLik(m0.glm))

## Error in logLik(m0.glm): Objekt 'm0.glm' nicht gefunden

aic.m0.glmer; aic.glm

## Error in eval(expr, envir, enclos): Objekt 'aic.m0.glmer' nicht gefunden

## Error in eval(expr, envir, enclos): Objekt 'aic.glm' nicht gefunden

### Model fitting

Extract interactions

# find all 2-way interactions  
library(utils)  
#colnames(pphi)  
# define variables included in interactions  
vars <- c("Variant", "Date", "Gender", "Age" , "Function", "Frequency",   
 "Priming", "Gradability", "SemanticCategory", "Emotionality")  
# 2-way interactions (involving Variant)  
intac\_2way <- t(combn(vars, 2))  
intac\_2way <- intac\_2way[which(intac\_2way[,1] == "Variant"),]  
int2w <- paste(intac\_2way[,1], intac\_2way[,2], sep = ":")  
# 3-way interactions (involving Variant)  
intac\_3way <- t(combn(vars, 3))  
intac\_3way <- intac\_3way[which(intac\_3way[,1] == "Variant"),]  
int3w <- paste(intac\_3way[,1], intac\_3way[,2], intac\_3way[,3], sep = ":")  
# combine vars, int2w and int3w  
predictors <- paste(vars, int2w, int3w, sep = " + ", collapse = " + ")  
# inspect  
predictors

## [1] "Variant + Variant:Date + Variant:Date:Gender + Date + Variant:Gender + Variant:Date:Age + Gender + Variant:Age + Variant:Date:Function + Age + Variant:Function + Variant:Date:Frequency + Function + Variant:Frequency + Variant:Date:Priming + Frequency + Variant:Priming + Variant:Date:Gradability + Priming + Variant:Gradability + Variant:Date:SemanticCategory + Gradability + Variant:SemanticCategory + Variant:Date:Emotionality + SemanticCategory + Variant:Emotionality + Variant:Gender:Age + Emotionality + Variant:Date + Variant:Gender:Function + Variant + Variant:Gender + Variant:Gender:Frequency + Date + Variant:Age + Variant:Gender:Priming + Gender + Variant:Function + Variant:Gender:Gradability + Age + Variant:Frequency + Variant:Gender:SemanticCategory + Function + Variant:Priming + Variant:Gender:Emotionality + Frequency + Variant:Gradability + Variant:Age:Function + Priming + Variant:SemanticCategory + Variant:Age:Frequency + Gradability + Variant:Emotionality + Variant:Age:Priming + SemanticCategory + Variant:Date + Variant:Age:Gradability + Emotionality + Variant:Gender + Variant:Age:SemanticCategory + Variant + Variant:Age + Variant:Age:Emotionality + Date + Variant:Function + Variant:Function:Frequency + Gender + Variant:Frequency + Variant:Function:Priming + Age + Variant:Priming + Variant:Function:Gradability + Function + Variant:Gradability + Variant:Function:SemanticCategory + Frequency + Variant:SemanticCategory + Variant:Function:Emotionality + Priming + Variant:Emotionality + Variant:Frequency:Priming + Gradability + Variant:Date + Variant:Frequency:Gradability + SemanticCategory + Variant:Gender + Variant:Frequency:SemanticCategory + Emotionality + Variant:Age + Variant:Frequency:Emotionality + Variant + Variant:Function + Variant:Priming:Gradability + Date + Variant:Frequency + Variant:Priming:SemanticCategory + Gender + Variant:Priming + Variant:Priming:Emotionality + Age + Variant:Gradability + Variant:Gradability:SemanticCategory + Function + Variant:SemanticCategory + Variant:Gradability:Emotionality + Frequency + Variant:Emotionality + Variant:SemanticCategory:Emotionality"

Create formula

Formula <- formula(paste("Probability ~ ", "(1|Adjective/Variant) + ", predictors))  
# full model with all variables  
fm <- lmer(Formula, data = pphi)

## Error in lmer(Formula, data = pphi): konnte Funktion "lmer" nicht finden

# inspect  
#fm

AIC-based automated model fitting

library(lmerTest)

## Error in library(lmerTest): there is no package called 'lmerTest'

fixmodel <- lm(formula(fm, fixed.only = TRUE),  
 data = eval(getCall(fm)$data))

## Error in formula(fm, fixed.only = TRUE): Objekt 'fm' nicht gefunden

lmerTest::step(fixmodel, k = log(nrow(pphi)))

## Error in loadNamespace(name): there is no package called 'lmerTest'

Compare final minimal adequate model to intercept-only base-line model

mf.glmer <- glmer(Probability ~ (1 | Adjective/Variant) + Variant + Date + Gender + Age + Function +   
 Priming + Gradability + Emotionality + Variant:Date + Variant:Gender +   
 Variant:Age + Variant:Function + Variant:Priming + Variant:Gradability +   
 Variant:Emotionality + Variant:Date:Gender + Variant:Date:Function +   
 Variant:Gender:Function + Variant:Gender:Priming, family = "gaussian", data = pphi)

## Error in glmer(Probability ~ (1 | Adjective/Variant) + Variant + Date + : konnte Funktion "glmer" nicht finden

# compare models   
anova(mf.glmer, m0.glmer, test = "Chi")

## Error in anova(mf.glmer, m0.glmer, test = "Chi"): Objekt 'mf.glmer' nicht gefunden

Anova(mf.glmer, type = "III", test = "Chi")

## Error in Anova(mf.glmer, type = "III", test = "Chi"): konnte Funktion "Anova" nicht finden

### Tabulate regression results

library(sjPlot)

## Error in library(sjPlot): there is no package called 'sjPlot'

tab\_model(mf.glmer, p.val = "kr", show.df = TRUE)

## Error in tab\_model(mf.glmer, p.val = "kr", show.df = TRUE): konnte Funktion "tab\_model" nicht finden

### Visualize effects

Plot effects

library(effects)

## Error in library(effects): there is no package called 'effects'

png("D:\\Uni\\Projekte\\02-Intensification\\AmpHKPIE\\images/effectsfinalmodel.png", width = 960, height = 480)   
plot(allEffects(mf.glmer), type="response", ylim=c(0,1), grid=TRUE,   
 lines = list(col="black", lty = 1, confint=list(style="bars", col = "grey80")), ylab = "Probability")

## Error in allEffects(mf.glmer): konnte Funktion "allEffects" nicht finden

dev.off()

## png   
## 2

#plot(allEffects(mf.glmer), type="response", ylim=c(0,1), grid=TRUE,   
# lines = list(col="black", lty = 1, confint=list(style="bars", col = "grey80")), ylab = "Probability")

### Visualize individual effects

Custom plot

p\_age\_var <- plot\_model(mf.glmer, type = "pred", terms = c("Age", "Variant"))

## Error in plot\_model(mf.glmer, type = "pred", terms = c("Age", "Variant")): konnte Funktion "plot\_model" nicht finden

#p\_age\_var

Publishable plot

ggplot(pphi, aes(Age, Probability)) +  
 facet\_grid(cols = vars(Variant)) +  
 stat\_summary(fun.y = mean, geom = "point", size = 1.5, color = "gray20") +  
 stat\_summary(fun.data = mean\_cl\_boot, geom = "errorbar",   
 width = 0.2, size = .5, color = "gray20") +  
 coord\_cartesian(ylim = c(-.1, 1)) +  
 theme\_set(theme\_bw(base\_size = 12)) +  
 theme(legend.position="top",  
 panel.grid.major = element\_blank(),   
 panel.grid.minor = element\_blank(),   
 axis.text.x = element\_text(size=10)) +  
 labs(x = "Age", y = "Predicted probability") +  
 scale\_x\_discrete(limits = rev(levels(pphi$Age))) +  
 ggsave("D:\\Uni\\Projekte\\02-Intensification\\AmpHKPIE\\images/Pred\_AgeVariant.png",  
 height = 5, width = 6, dpi = 320)

## Error in ggplot(pphi, aes(Age, Probability)): konnte Funktion "ggplot" nicht finden

Custom plot

p\_emo\_var <- plot\_model(mf.glmer, type = "pred", terms = c("Emotionality", "Variant"))

## Error in plot\_model(mf.glmer, type = "pred", terms = c("Emotionality", : konnte Funktion "plot\_model" nicht finden

#p\_emo\_var

Publishable plot

ggplot(pphi, aes(Emotionality, Probability)) +  
 facet\_grid(cols = vars(Variant)) +  
 stat\_summary(fun.y = mean, geom = "point", size = 1.5, color = "gray20") +  
 stat\_summary(fun.data = mean\_cl\_boot, geom = "errorbar",   
 width = 0.2, size = .5, color = "gray20") +  
 coord\_cartesian(ylim = c(-.1, 1)) +  
 theme\_set(theme\_bw(base\_size = 12)) +  
 theme(legend.position="top",  
 panel.grid.major = element\_blank(),   
 panel.grid.minor = element\_blank(),   
 axis.text.x = element\_text(size=10, angle = 90)) +  
 labs(x = "Emotionality", y = "Predicted probability") +  
 ggsave("D:\\Uni\\Projekte\\02-Intensification\\AmpHKPIE\\images/Pred\_EmoVariant.png",  
 height = 5, width = 6, dpi = 320)

## Error in ggplot(pphi, aes(Emotionality, Probability)): konnte Funktion "ggplot" nicht finden

Custom plot

p\_grad\_var <- plot\_model(mf.glmer, type = "pred", terms = c("Gradability", "Variant"))

## Error in plot\_model(mf.glmer, type = "pred", terms = c("Gradability", : konnte Funktion "plot\_model" nicht finden

#p\_grad\_var

Publishable plot

# add prediction  
pphi$Prediction <- predict(mf.glmer, pphi, type="response")

## Error in predict(mf.glmer, pphi, type = "response"): Objekt 'mf.glmer' nicht gefunden

# start plotting  
ggplot(pphi, aes(Gradability, Prediction, group = Variant, color = Variant, linetype = Variant)) +  
 geom\_smooth(se = F, method = "lm") +  
 coord\_cartesian(ylim = c(0, 1)) +  
 theme\_set(theme\_bw(base\_size = 12)) +  
 theme(legend.position="top",   
 panel.grid.major = element\_blank(),   
 panel.grid.minor = element\_blank(),   
 axis.text.x = element\_text(size=10)) +  
 labs(x = "Gradability", y = "Predicted probability") +  
 scale\_color\_manual(breaks = c("other", "really", "so", "very"),  
 values = c("gray80", "gray60", "gray40", "gray20")) +  
 scale\_linetype\_manual(breaks = c("other", "really", "so", "very"),  
 values = c(4, 3, 2, 1)) +   
 ggsave("D:\\Uni\\Projekte\\02-Intensification\\AmpHKPIE\\images/Pred\_GradabilityVariant.png",  
 height = 4, width = 5, dpi = 320)

## Error in ggplot(pphi, aes(Gradability, Prediction, group = Variant, color = Variant, : konnte Funktion "ggplot" nicht finden

Custom plot

p\_date\_gender\_var <- plot\_model(mf.glmer, type = "pred", terms = c("Date", "Gender", "Variant"))

## Error in plot\_model(mf.glmer, type = "pred", terms = c("Date", "Gender", : konnte Funktion "plot\_model" nicht finden

#p\_date\_gender\_var

Publishable plot

ggplot(pphi, aes(Date, Probability, group = Gender, color = Gender)) +  
 facet\_grid(cols = vars(Variant)) +  
 stat\_summary(fun.y = mean, geom = "point", size = 1.5) +  
 stat\_summary(fun.data = mean\_cl\_boot, geom = "errorbar",   
 width = 0.2, size = .5) +  
 coord\_cartesian(ylim = c(-.1, 1)) +  
 theme\_set(theme\_bw(base\_size = 12)) +  
 theme(legend.position="top",  
 panel.grid.major = element\_blank(),   
 panel.grid.minor = element\_blank(),   
 axis.text.x = element\_text(size=10, angle = 90)) +  
 scale\_color\_manual(breaks = c("female", "male"),  
 values = c("gray60", "gray20")) +  
 labs(x = "Date", y = "Predicted probability") +  
 ggsave("D:\\Uni\\Projekte\\02-Intensification\\AmpHKPIE\\images/Pred\_DateGenderVariant.png",  
 height = 5, width = 6, dpi = 320)

## Error in ggplot(pphi, aes(Date, Probability, group = Gender, color = Gender)): konnte Funktion "ggplot" nicht finden

Custom plot

p\_date\_priming\_var <- plot\_model(mf.glmer, type = "pred", terms = c("Date", "Priming", "Variant"))

## Error in plot\_model(mf.glmer, type = "pred", terms = c("Date", "Priming", : konnte Funktion "plot\_model" nicht finden

#p\_date\_priming\_var

Publishable plot

ggplot(pphi, aes(Date, Probability, group = Priming, color = Priming)) +  
 facet\_grid(cols = vars(Variant)) +  
 stat\_summary(fun.y = mean, geom = "point", size = 1.5) +  
 stat\_summary(fun.data = mean\_cl\_boot, geom = "errorbar",   
 width = 0.2, size = .5) +  
 coord\_cartesian(ylim = c(-.1, 1)) +  
 theme\_set(theme\_bw(base\_size = 12)) +  
 theme(legend.position="top",  
 panel.grid.major = element\_blank(),   
 panel.grid.minor = element\_blank(),   
 axis.text.x = element\_text(size=10, angle = 90)) +  
 scale\_color\_manual(breaks = c("NotPrimed", "Primed"),  
 values = c("gray60", "gray20")) +  
 labs(x = "Date", y = "Predicted probability") +  
 ggsave("D:\\Uni\\Projekte\\02-Intensification\\AmpHKPIE\\images/Pred\_DatePrimingVariant.png",  
 height = 5, width = 6, dpi = 320)

## Error in ggplot(pphi, aes(Date, Probability, group = Priming, color = Priming)): konnte Funktion "ggplot" nicht finden

Custom plot

Custom plot

p\_gender\_priming\_var <- plot\_model(mf.glmer, type = "pred", terms = c("Priming", "Gender", "Variant"))

## Error in plot\_model(mf.glmer, type = "pred", terms = c("Priming", "Gender", : konnte Funktion "plot\_model" nicht finden

#p\_gender\_priming\_var

Publishable plot

ggplot(pphi, aes(Priming, Probability, group = Gender, color = Gender)) +  
 facet\_grid(cols = vars(Variant)) +  
 stat\_summary(fun.y = mean, geom = "point", size = 1.5) +  
 stat\_summary(fun.data = mean\_cl\_boot, geom = "errorbar",   
 width = 0.2, size = .5) +  
 coord\_cartesian(ylim = c(-.1, 1)) +  
 theme\_set(theme\_bw(base\_size = 12)) +  
 theme(legend.position="top",  
 panel.grid.major = element\_blank(),   
 panel.grid.minor = element\_blank(),   
 axis.text.x = element\_text(size=10, angle = 90)) +  
 scale\_color\_manual(breaks = c("female", "male"),  
 values = c("gray60", "gray20")) +  
 labs(x = "Date", y = "Predicted probability") +  
 ggsave("D:\\Uni\\Projekte\\02-Intensification\\AmpHKPIE\\images/Pred\_PrimingGenderVariant.png",  
 height = 5, width = 6, dpi = 320)

## Error in ggplot(pphi, aes(Priming, Probability, group = Gender, color = Gender)): konnte Funktion "ggplot" nicht finden

Custom plot

p\_age\_function\_var <- plot\_model(mf.glmer, type = "pred", terms = c("Age", "Function", "Variant"))

## Error in plot\_model(mf.glmer, type = "pred", terms = c("Age", "Function", : konnte Funktion "plot\_model" nicht finden

#p\_age\_function\_var

Publishable plot

# start plotting  
ggplot(pphi, aes(Age, Prediction, group = Variant, color = Variant, linetype = Variant)) +  
 facet\_grid(cols = vars(Function)) +  
 geom\_smooth(se = F, method = "lm") +  
 coord\_cartesian(ylim = c(-0.1, 1)) +  
 theme\_set(theme\_bw(base\_size = 12)) +  
 theme(legend.position="top",   
 panel.grid.major = element\_blank(),   
 panel.grid.minor = element\_blank(),   
 axis.text.x = element\_text(size=10)) +  
 labs(x = "Age", y = "Predicted probability") +  
 scale\_color\_manual(breaks = c("other", "really", "so", "very"),  
 values = c("gray80", "gray60", "gray40", "gray20")) +  
 scale\_linetype\_manual(breaks = c("other", "really", "so", "very"),  
 values = c(4, 3, 2, 1)) +   
 scale\_x\_discrete(limits = rev(levels(pphi$Age))) +  
 ggsave("D:\\Uni\\Projekte\\02-Intensification\\AmpHKPIE\\images/Pred\_AgeFunctionVariant.png",  
 height = 5, width = 6, dpi = 320)

## Error in ggplot(pphi, aes(Age, Prediction, group = Variant, color = Variant, : konnte Funktion "ggplot" nicht finden

### Visualize random effects

randomtb <- ranef(mf.glmer)$Adjective

## Error in ranef(mf.glmer): konnte Funktion "ranef" nicht finden

rndmlngtb <- data.frame(rownames(randomtb), randomtb)

## Error in rownames(randomtb): Objekt 'randomtb' nicht gefunden

colnames(rndmlngtb) <- c("Adjective", "Intercept")

## Error in colnames(rndmlngtb) <- c("Adjective", "Intercept"): Objekt 'rndmlngtb' nicht gefunden

rndmlngtb <- rndmlngtb[order(rndmlngtb$Intercept, decreasing = T),]

## Error in eval(expr, envir, enclos): Objekt 'rndmlngtb' nicht gefunden

rndmlngtb

## Error in eval(expr, envir, enclos): Objekt 'rndmlngtb' nicht gefunden

summary(rndmlngtb$Intercept)

## Error in summary(rndmlngtb$Intercept): Objekt 'rndmlngtb' nicht gefunden

Publishable plot

ggplot(rndmlngtb, aes(Adjective, Intercept)) +  
 geom\_point(aes(reorder(Adjective, -Intercept, fun = Intercept), y=Intercept)) +  
 coord\_cartesian(ylim = c(-1, 2)) +  
 theme\_set(theme\_bw(base\_size = 12)) +  
 theme(legend.position="none",  
 panel.grid.major = element\_blank(),   
 panel.grid.minor = element\_blank(),   
 axis.text.x = element\_text(size=10, angle=90)) +  
 labs(x = "Adjective type", y = "Adjustment to Intercept") +  
 ggsave("D:\\Uni\\Projekte\\02-Intensification\\AmpHKPIE\\images/Pred\_Adjectives.png",  
 height = 4, width = 5, dpi = 320)

## Error in ggplot(rndmlngtb, aes(Adjective, Intercept)): konnte Funktion "ggplot" nicht finden

## Power analysis

# load package  
library(simr)

## Error in library(simr): there is no package called 'simr'

# rename model  
gm1 <- mf.glmer

## Error in eval(expr, envir, enclos): Objekt 'mf.glmer' nicht gefunden

# set seed for replicability  
#set.seed(2020011401)  
# calculate power for ConversationType  
#powerSim(gm1, fixed("ConversationType", "lr"), nsim=100)

Extract fixed effecst estimates

estimatesfixedeffects <- fixef(mf.glmer)

## Error in fixef(mf.glmer): konnte Funktion "fixef" nicht finden

exp(estimatesfixedeffects)

## Error in eval(expr, envir, enclos): Objekt 'estimatesfixedeffects' nicht gefunden

Change effect size of Variant[really] to make it “medium” (odds ratio = 3.47 or Cohen’s d of .5, estimate = 1.245)

fixef(gm1)["Variantreally"] <- 0.519

## Error in fixef(gm1)["Variantreally"] <- 0.519: Objekt 'gm1' nicht gefunden

estimatesfixedeffects <- fixef(gm1)

## Error in fixef(gm1): konnte Funktion "fixef" nicht finden

exp(estimatesfixedeffects)

## Error in eval(expr, envir, enclos): Objekt 'estimatesfixedeffects' nicht gefunden

Percent accuracy of the model to detect a medium effect with a likelihood higher than 80 percent.

# set seed for replicability  
set.seed(2020011405)  
powerSim(gm1, fixed("Variantreally", "z"), nsim=100)

## Error in powerSim(gm1, fixed("Variantreally", "z"), nsim = 100): konnte Funktion "powerSim" nicht finden

Extract session information

sessionInfo()

## R version 4.0.2 (2020-06-22)  
## Platform: x86\_64-w64-mingw32/x64 (64-bit)  
## Running under: Windows 10 x64 (build 18362)  
##   
## Matrix products: default  
##   
## locale:  
## [1] LC\_COLLATE=German\_Germany.1252 LC\_CTYPE=German\_Germany.1252   
## [3] LC\_MONETARY=German\_Germany.1252 LC\_NUMERIC=C   
## [5] LC\_TIME=German\_Germany.1252   
##   
## attached base packages:  
## [1] stats graphics grDevices utils datasets methods base   
##   
## loaded via a namespace (and not attached):  
## [1] compiler\_4.0.2 magrittr\_1.5 htmltools\_0.5.0 tools\_4.0.2   
## [5] yaml\_2.2.1 stringi\_1.4.6 rmarkdown\_2.3 knitr\_1.29   
## [9] stringr\_1.4.0 xfun\_0.16 digest\_0.6.25 packrat\_0.5.0   
## [13] rlang\_0.4.7 evaluate\_0.14

=========================THE END=========================