Metaphor usage patterns

Peer Christensen

10/10/2017

library(data.table)  
library(ggplot2)  
library(viridis)

## Loading required package: viridisLite

library(Hmisc)

## Loading required package: lattice

## Loading required package: survival

## Loading required package: Formula

##   
## Attaching package: 'Hmisc'

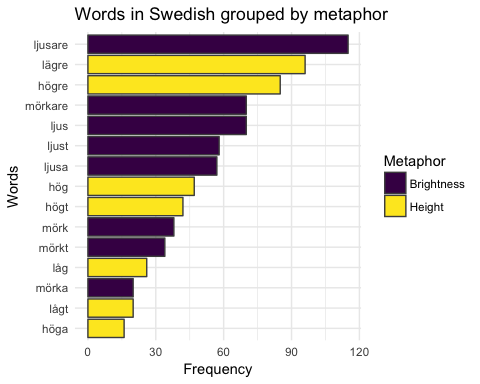
## The following objects are masked from 'package:base':  
##   
## format.pval, round.POSIXt, trunc.POSIXt, units

library(plyr)

##   
## Attaching package: 'plyr'

## The following objects are masked from 'package:Hmisc':  
##   
## is.discrete, summarize

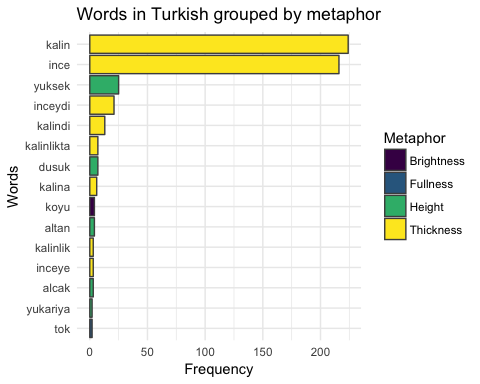
setwd("/Users/peerchristensen/Desktop")  
df\_full= read.csv2("all\_data\_CLEAN.csv",na.strings = c(""),stringsAsFactors = F)  
#remove some columns  
df=df\_full[,-c(1,3,13,14,15,16,17,18)]  
df=df[!duplicated(df), ]  
  
########## 1. WORDS ########  
  
#### 1.1 Words by metaphor ###  
wordsMeta=df[,c(5,6,11)]  
wordsMeta=wordsMeta[wordsMeta$Scope!="none" & wordsMeta$Scope!="NA",]  
wordsMeta$Scope[wordsMeta$Words=="tok"]="Fullness"  
wordsMeta=data.table(na.omit(wordsMeta))  
  
#Swedish  
wordsMetaSwe=wordsMeta[wordsMeta$Language=="Swedish"]  
wordsMetaSwe[,Freq := .N, by=Words]  
wordsMetaSwe=unique(wordsMetaSwe)  
wordsMetaSwe=wordsMetaSwe[order(-Freq)]  
wordsMetaSwe=wordsMetaSwe[1:15,]  
  
#wordsMetaSwe=data.frame(sort(table(wordsMetaSwe$Scope), decreasing=T))  
  
wmSwe=ggplot(wordsMetaSwe,aes(reorder(Words,Freq),Freq)) +   
 geom\_bar(stat = "identity",aes(fill=Scope),colour="gray30") +  
 scale\_fill\_viridis("Metaphor",discrete = TRUE, option = "D") +  
 theme(axis.text.x = element\_text(angle = 90, hjust = 1)) +  
 ggtitle("Words in Swedish grouped by metaphor") +  
 xlab("Words") +  
 ylab("Frequency") +  
 coord\_flip() +  
 theme\_minimal()  
wmSwe



#Turkish  
wordsMetaTur=wordsMeta[wordsMeta$Language=="Turkish"]  
wordsMetaTur$Scope[wordsMetaTur=="tok"]="Fullness"

## Warning in `[<-.data.table`(x, j = name, value = value): Supplied 1794  
## items to be assigned to 598 items of column 'Scope' (1196 unused)

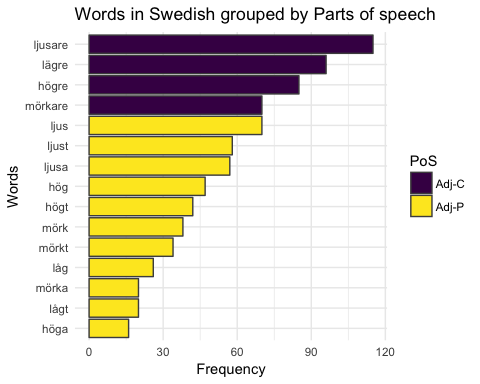
wordsMetaTur[,Freq := .N, by=Words]  
wordsMetaTur=unique(wordsMetaTur)  
wordsMetaTur=wordsMetaTur[order(-Freq)]  
wordsMetaTur=wordsMetaTur[1:15,]  
wmTur=ggplot(wordsMetaTur,aes(reorder(Words,Freq),Freq)) +   
 geom\_bar(stat = "identity",aes(fill=Scope),colour="gray30") +  
 scale\_fill\_viridis("Metaphor",discrete = TRUE, option = "D") +  
 theme(axis.text.x = element\_text(angle = 90, hjust = 1)) +  
 ggtitle("Words in Turkish grouped by metaphor") +  
 xlab("Words") +  
 ylab("Frequency") +  
 coord\_flip() +  
 theme\_minimal()  
wmTur



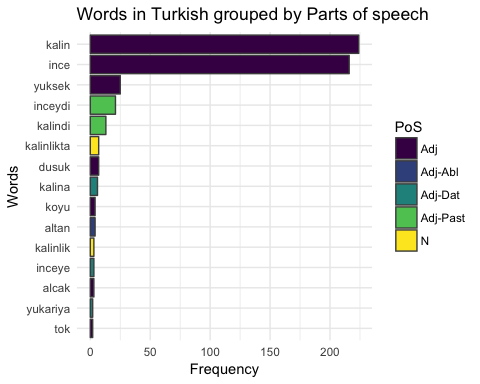
#### 1.2 Words by PoS ###  
wordsPoS=df[,c(6,11)]  
wordsPoS$PoS=revalue(wordsPoS$Words, c("ljusare"="Adj-C","lägre"="Adj-C",  
 "högre"="Adj-C","mörkare"="Adj-C",  
 "ljus"="Adj-P","ljust"="Adj-P","ljusa"="Adj-P",  
 "mörk"="Adj-P","mörka"="Adj-P","mörkt"="Adj-P",  
 "hög"="Adj-P","högt"="Adj-P","höga"="Adj-P",  
 "låg"="Adj-P","lågt"="Adj-P","högre upp"="ADV",  
 "högt upp"="ADV","upp"="ADV","över"="PREP",  
 "under"="PREP","djup"="Adj-P","första"="REF",  
 "andra"="REF","ner"="ADV","kalin"="Adj","ince"="Adj","yuksek"="Adj",  
 "inceydi"="Adj-Past","kalindi"="Adj-Past","kalinlikta"="N",  
 "dusuk"="Adj","kalina"="Adj-Dat","koyu"="Adj","altan"="Adj-Abl",  
 "alcak"="Adj","kalinlik"="N","inceye"="Adj-Dat","tok"="Adj","yukariya"="Adj-Dat"))

## The following `from` values were not present in `x`: över, första

wordsPoS=data.table(na.omit(wordsPoS))  
  
#Swedish  
wordsPoSSwe=wordsPoS[wordsPoS$Language=="Swedish"]  
wordsPoSSwe[,Freq := .N, by=Words]  
wordsPoSSwe=unique(wordsPoSSwe)  
wordsPoSSwe=wordsPoSSwe[order(-Freq)]  
wordsPoSSwe=wordsPoSSwe[1:15,]  
  
PoSSwe=ggplot(wordsPoSSwe,aes(reorder(Words,Freq),Freq)) +   
 geom\_bar(stat = "identity",aes(fill=PoS),colour="gray30") +  
 scale\_fill\_viridis("PoS",discrete = TRUE, option = "D") +  
 theme(axis.text.x = element\_text(angle = 90, hjust = 1)) +  
 ggtitle("Words in Swedish grouped by Parts of speech") +  
 xlab("Words") +  
 ylab("Frequency") +  
 coord\_flip() +  
 theme\_minimal()  
PoSSwe



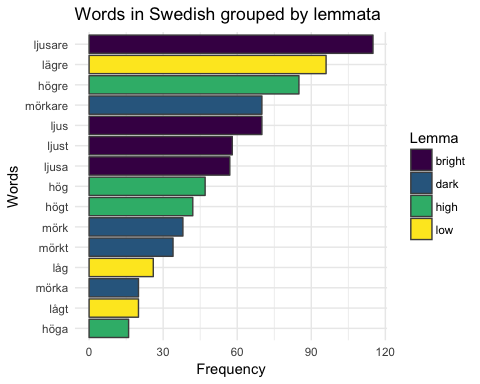
#Turkish  
wordsPoSTur=wordsPoS[wordsPoS$Language=="Turkish"]  
wordsPoSTur[,Freq := .N, by=Words]  
wordsPoSTur=unique(wordsPoSTur)  
wordsPoSTur=wordsPoSTur[order(-Freq)]  
wordsPoSTur=wordsPoSTur[1:15,]  
  
PoSTur=ggplot(wordsPoSTur,aes(reorder(Words,Freq),Freq)) +   
 geom\_bar(stat = "identity",aes(fill=PoS),colour="gray30") +  
 scale\_fill\_viridis("PoS",discrete = TRUE, option = "D") +  
 theme(axis.text.x = element\_text(angle = 90, hjust = 1)) +  
 ggtitle("Words in Turkish grouped by Parts of speech") +  
 xlab("Words") +  
 ylab("Frequency") +  
 coord\_flip() +  
 theme\_minimal()  
PoSTur



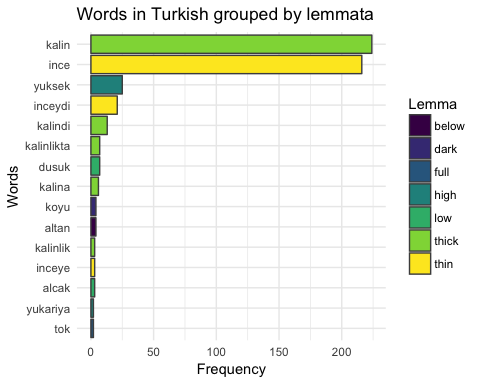
### 1.3a Words by lemmata ###  
wordsLem=df[,c(6,11)]  
wordsLem$Lem=revalue(wordsPoS$Words, c("ljusare"="bright","lägre"="low","djupa"="deep",  
 "högre"="high","mörkare"="dark",  
 "ljus"="bright","ljust"="bright","ljusa"="bright",  
 "mörk"="dark","mörka"="dark","mörkt"="dark",  
 "hög"="high","höjd"="high","högt"="high","höga"="high",  
 "låg"="low","lågt"="low","högre upp"="high up",  
 "högt upp"="high up","upp"="up","över"="over",  
 "under"="under","djup"="deep","första"="first",  
 "andra"="other","ner"="down","kalin"="thick","ince"="thin","yuksek"="high",  
 "inceydi"="thin","kalindi"="thick","kalinlikta"="thick",  
 "dusuk"="low","kalina"="thick","koyu"="dark","altan"="below",  
 "alcak"="low","kalinlik"="thick","inceye"="thin","tok"="full","yukariya"="high"))

## The following `from` values were not present in `x`: över, första

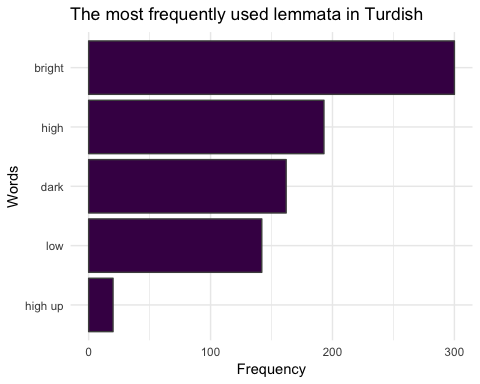
wordsLem=data.table(na.omit(wordsLem))  
  
#Swedish  
wordsLemSwe=wordsLem[wordsLem$Language=="Swedish"]  
wordsLemSwe[,Freq := .N, by=Words]  
wordsLemSwe=unique(wordsLemSwe)  
wordsLemSwe=wordsLemSwe[order(-Freq)]  
wordsLemSwe=wordsLemSwe[1:15,]  
  
LemSwe=ggplot(wordsLemSwe,aes(reorder(Words,Freq),Freq)) +   
 geom\_bar(stat = "identity",aes(fill=Lem),colour="gray30") +  
 scale\_fill\_viridis("Lemma",discrete = TRUE, option = "D") +  
 theme(axis.text.x = element\_text(angle = 90, hjust = 1)) +  
 ggtitle("Words in Swedish grouped by lemmata") +  
 xlab("Words") +  
 ylab("Frequency") +  
 coord\_flip() +  
 theme\_minimal()  
LemSwe



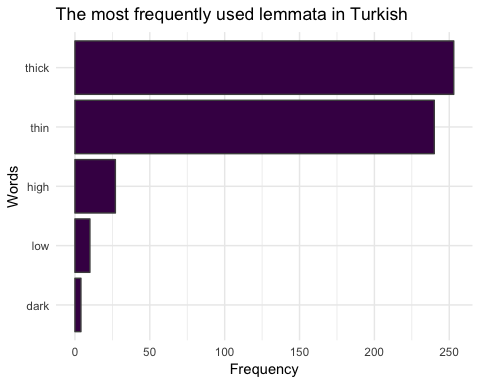
#Turkish  
wordsLemTur=wordsLem[wordsLem$Language=="Turkish"]  
wordsLemTur[,Freq := .N, by=Words]  
wordsLemTur=unique(wordsLemTur)  
wordsLemTur=wordsLemTur[order(-Freq)]  
wordsLemTur=wordsLemTur[1:15,]  
  
LemTur=ggplot(wordsLemTur,aes(reorder(Words,Freq),Freq)) +   
 geom\_bar(stat = "identity",aes(fill=Lem),colour="gray30") +  
 scale\_fill\_viridis("Lemma",discrete = TRUE, option = "D") +  
 theme(axis.text.x = element\_text(angle = 90, hjust = 1)) +  
 ggtitle("Words in Turkish grouped by lemmata") +  
 xlab("Words") +  
 ylab("Frequency") +  
 coord\_flip() +  
 theme\_minimal()  
LemTur



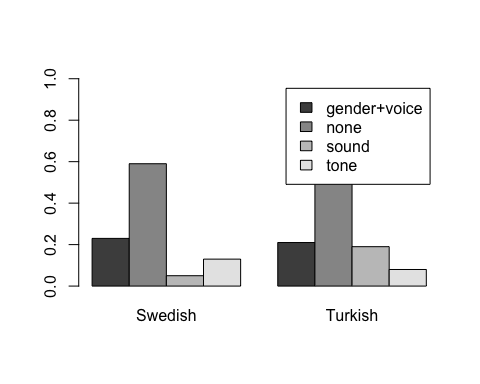
### 1.3b Lemmata ###  
wordsLem2=wordsLem[,-1]  
  
#Swedish  
wordsLemSwe2=wordsLem2[wordsLem2$Language=="Swedish"]  
wordsLemSwe2[,Freq := .N, by=Lem]  
wordsLemSwe2=unique(wordsLemSwe2)  
wordsLemSwe2=wordsLemSwe2[order(-Freq)]  
wordsLemSwe2=wordsLemSwe2[1:5,]  
  
LemSwe2=ggplot(wordsLemSwe2,aes(reorder(Lem,Freq),Freq)) +   
 geom\_bar(stat = "identity",fill="#440154FF",colour="gray30") +  
 scale\_fill\_viridis("Lemma",discrete = TRUE, option = "D") +  
 theme(axis.text.x = element\_text(angle = 90, hjust = 1)) +  
 ggtitle("The most frequently used lemmata in Turdish") +  
 xlab("Words") +  
 ylab("Frequency") +  
 coord\_flip() +  
 theme\_minimal()  
LemSwe2



#Turkish  
wordsLemTur2=wordsLem2[wordsLem2$Language=="Turkish"]  
wordsLemTur2[,Freq := .N, by=Lem]  
wordsLemTur2=unique(wordsLemTur2)  
wordsLemTur2=wordsLemTur2[order(-Freq)]  
wordsLemTur2=wordsLemTur2[1:5,]  
  
LemTur2=ggplot(wordsLemTur2,aes(reorder(Lem,Freq),Freq)) +   
 geom\_bar(stat = "identity",fill="#440154FF",colour="gray30") +  
 scale\_fill\_viridis("Lemma",discrete = TRUE, option = "D") +  
 theme(axis.text.x = element\_text(angle = 90, hjust = 1)) +  
 ggtitle("The most frequently used lemmata in Turkish") +  
 xlab("Words") +  
 ylab("Frequency") +  
 coord\_flip() +  
 theme\_minimal()  
LemTur2



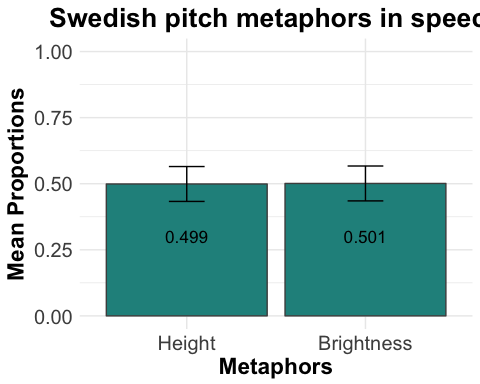
### 1.4 Subject ####  
dfSubj=df[df$Subject!="clip" & df$Subject!="instrument"&  
 df$Subject!="NA" & df$Subject!="part",]  
dfSubj$Subject[dfSubj$Subject=="voice"|dfSubj$Subject=="gender"]="gender+voice"  
Subjects=round(prop.table(table(dfSubj$Subject,dfSubj$Language),2),2)  
barplot(Subjects,beside=T,legend=T,ylim=c(0,1))



########## 2. METAPHORS ########  
  
#### 2.1 Metaphor usage ####  
  
#Swedish  
dfS=df[df$Language=="Swedish",]  
dfS=ftable(dfS$File,dfS$Scope)  
wtS=rowSums(dfS)  
dfS=round(prop.table(as.matrix(dfS),1),2)  
dfS=data.frame(dfS,wtS)  
dfS$Language="Swedish"  
  
dfS=ddply(dfS,"Language",summarise,  
 H= weighted.mean(Height,wtS),   
 B=weighted.mean(Brightness,wtS),  
 Hse=sqrt(wtd.var(Height,wtS))/sqrt(nrow(dfS)),  
 Bse=sqrt(wtd.var(Brightness,wtS))/sqrt(nrow(dfS)))  
  
dfS <- transform(dfS, Hlower=H-Hse, Hupper=H+Hse,Blower=B-Bse,  
 Bupper=B+Bse)  
  
dfS=melt(dfS,Language=Language)

## Using Language as id variables

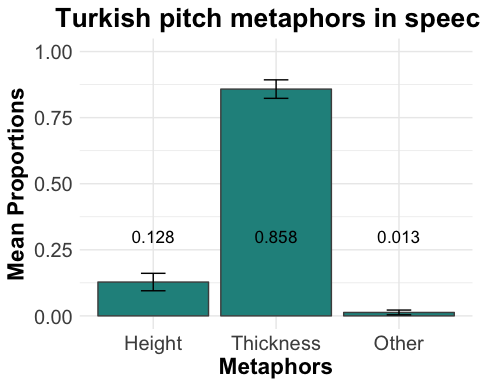
dfS=data.frame(cbind(dfS[1:2,],  
 dfS[3:4,2:3]))  
dfS[,c(3,5)]=round(dfS[,c(3,5)],3)  
  
BarSwe <- ggplot(dfS, aes(x=variable, y=value, fill=Language)) +   
 geom\_bar(position=position\_dodge(),stat="identity",fill="#21908CFF" ,colour="gray30") +  
 geom\_errorbar(aes(ymin=value-value.1, ymax=value+value.1), width=.2,  
 position=position\_dodge(.9)) +  
 geom\_text(aes(label = value, y = 0.3, size = 3)) +  
 scale\_y\_continuous(name="Mean Proportions", limits=c(0, 1)) +  
 scale\_x\_discrete(breaks=c("H","B"),  
 labels=c("Height", "Brightness")) +  
 xlab("Metaphors") +  
 ggtitle("Swedish pitch metaphors in speech") +  
 theme\_minimal() +  
 theme(plot.title=element\_text(lineheight=0.8,hjust=0.5,face="bold", size=20),  
 axis.title.x=element\_text(face="bold",size=17),   
 axis.text.x=element\_text(size=15),  
 axis.title.y=element\_text(face="bold",size=17),  
 axis.text.y=element\_text(size=15),  
 legend.position = "none")  
BarSwe



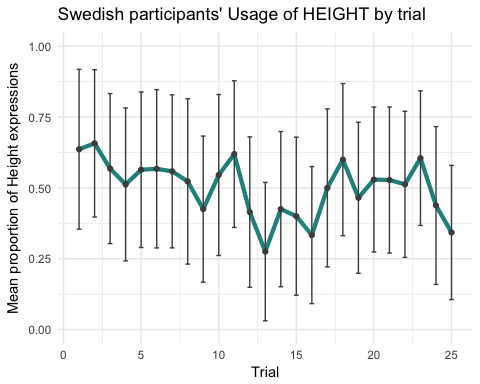
#Turkish  
dfT=df[df$Language=="Turkish",]  
#we collapse "Brightness" and "Other"  
dfT$Scope[dfT$Scope=="Brightness"]="Other"  
dfT=ftable(dfT$File,dfT$Scope)  
wtT=rowSums(dfT)  
dfT=round(prop.table(as.matrix(dfT),1),2)  
dfT=data.frame(dfT,wtT)  
dfT$Language="Turkish"  
  
dfT=ddply(dfT,"Language",summarise,  
 H= weighted.mean(Height,wtT),   
 Th=weighted.mean(Thickness,wtT),  
 O=weighted.mean(Other,wtT),  
 Hse=sqrt(wtd.var(Height,wtT))/sqrt(nrow(dfT)),  
 Thse=sqrt(wtd.var(Thickness,wtT))/sqrt(nrow(dfT)),  
 Ose=sqrt(wtd.var(Other,wtT))/sqrt(nrow(dfT)))  
  
dfT <- transform(dfT, Hlower=H-Hse, Hupper=H+Hse,Thlower=Th-Thse,  
 Thupper=Th+Thse,Olower=O-Ose,Oupper=O+Ose)  
  
dfT=melt(dfT,Language=Language)

## Using Language as id variables

dfT=data.frame(cbind(dfT[1:3,],  
 dfT[4:6,2:3]))  
  
dfT[,c(3,5)]=round(dfT[,c(3,5)],3)  
  
BarTur <- ggplot(dfT, aes(x=variable, y=value, fill=Language)) +   
 geom\_bar(position=position\_dodge(),stat="identity",fill="#21908CFF" ,colour="gray30") +  
 geom\_errorbar(aes(ymin=value-value.1, ymax=value+value.1), width=.2,  
 position=position\_dodge(.9)) +  
 geom\_text(aes(label = value, y = 0.3, size = 3)) +  
 scale\_y\_continuous(name="Mean Proportions", limits=c(0, 1)) +  
 scale\_x\_discrete(breaks=c("H","Th","O"),  
 labels=c("Height","Thickness","Other")) +  
 xlab("Metaphors") +  
 ggtitle("Turkish pitch metaphors in speech") +  
 theme\_minimal() +  
 theme(plot.title=element\_text(lineheight=0.8,hjust=0.5,face="bold", size=20),  
 axis.title.x=element\_text(face="bold",size=17),   
 axis.text.x=element\_text(size=15),  
 axis.title.y=element\_text(face="bold",size=17),  
 axis.text.y=element\_text(size=15),  
 legend.position = "none")  
BarTur



#### 2.2 Consistency over time ####  
#Swedish height and Turkish thickness, lineplot with error bars  
  
#Swedish  
conS=df[df$Language=="Swedish" & !df$Trial>25,]  
conS=conS[,-c(4,6,7,8,9,10,11)]  
conS=conS[,c(1,2,4)]  
wMeans=c()  
wSEs=c()  
for(i in seq(1,25)){  
 t=conS[conS$Trial==i,]  
 a=ftable(t,col.vars = "File")  
 b=colSums(a)  
 a=round(prop.table(as.matrix(a),2),2)  
 a=rbind(a,b)  
 wMean=weighted.mean(a[2,],a[3,])  
 wMeans=append(wMeans,wMean)  
 wSE=sqrt(wtd.var(a[2,],a[3,]))/sqrt(nrow(a))  
 wSEs=append(wSEs,wSE)  
}  
conS=data.frame(Trial=seq(1,25),wMeans,wSEs)  
  
consistencyS=ggplot(conS,aes(x=Trial, y=wMeans)) +  
 geom\_line(size = 1.5,color="#21908CFF") +  
 geom\_point(size=1.5,color="gray30") +  
 geom\_errorbar(aes(ymin=wMeans-wSEs,ymax=wMeans+wSEs),width=0.3,color="gray30") +  
 scale\_y\_continuous(name="Mean proportion of Height expressions", limits=c(0,1)) +  
 ggtitle("Swedish participants' Usage of HEIGHT by trial") +  
 ylab("Trial") +  
 theme\_minimal()  
consistencyS



#Turkish  
conT=df[df$Language=="Turkish" & !df$Trial>25,]  
conT=conT[,-c(4,6,7,8,9,10,11)]  
conT=conT[,c(1,2,4)]  
conT = conT[order(conT$Trial,conT$File),]   
wMeans=c()  
wSEs=c()  
for(i in seq(1,25)){  
 t=conT[conT$Trial==i,]  
 a=ftable(t,col.vars = "File")  
 b=colSums(a)  
 a=round(prop.table(as.matrix(a),2),2)  
 a=rbind(a,b)  
 wMean=weighted.mean(tail(a,2)[1,],tail(a,2)[2,])  
 wMeans=append(wMeans,wMean)  
 wSE=sqrt(wtd.var(tail(a,2)[1,],tail(a,2)[2,]))/sqrt(nrow(a))  
 wSEs=append(wSEs,wSE)  
}  
conT=data.frame(Trial=seq(1,25),wMeans,wSEs)  
  
consistencyT=ggplot(conT,aes(x=Trial, y=wMeans)) +  
 geom\_line(size = 1.5,color="#21908CFF") +  
 geom\_point(size=1.5,color="gray30") +  
 geom\_errorbar(aes(ymin=wMeans-wSEs,ymax=wMeans+wSEs),width=0.3,color="gray30") +  
 scale\_y\_continuous(name="Mean proportion of Height expressions", limits=c(0,1.1)) +  
 ggtitle("Turkish participants' Usage of THICKNESS by trial") +  
 ylab("Trial") +  
 theme\_minimal()  
consistencyT

