

Metaphor usage patterns

Peer Christensen

10/10/2017

```
# FINAL ANALYSIS MUSIC EXP DATA
# OCTOBER 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))

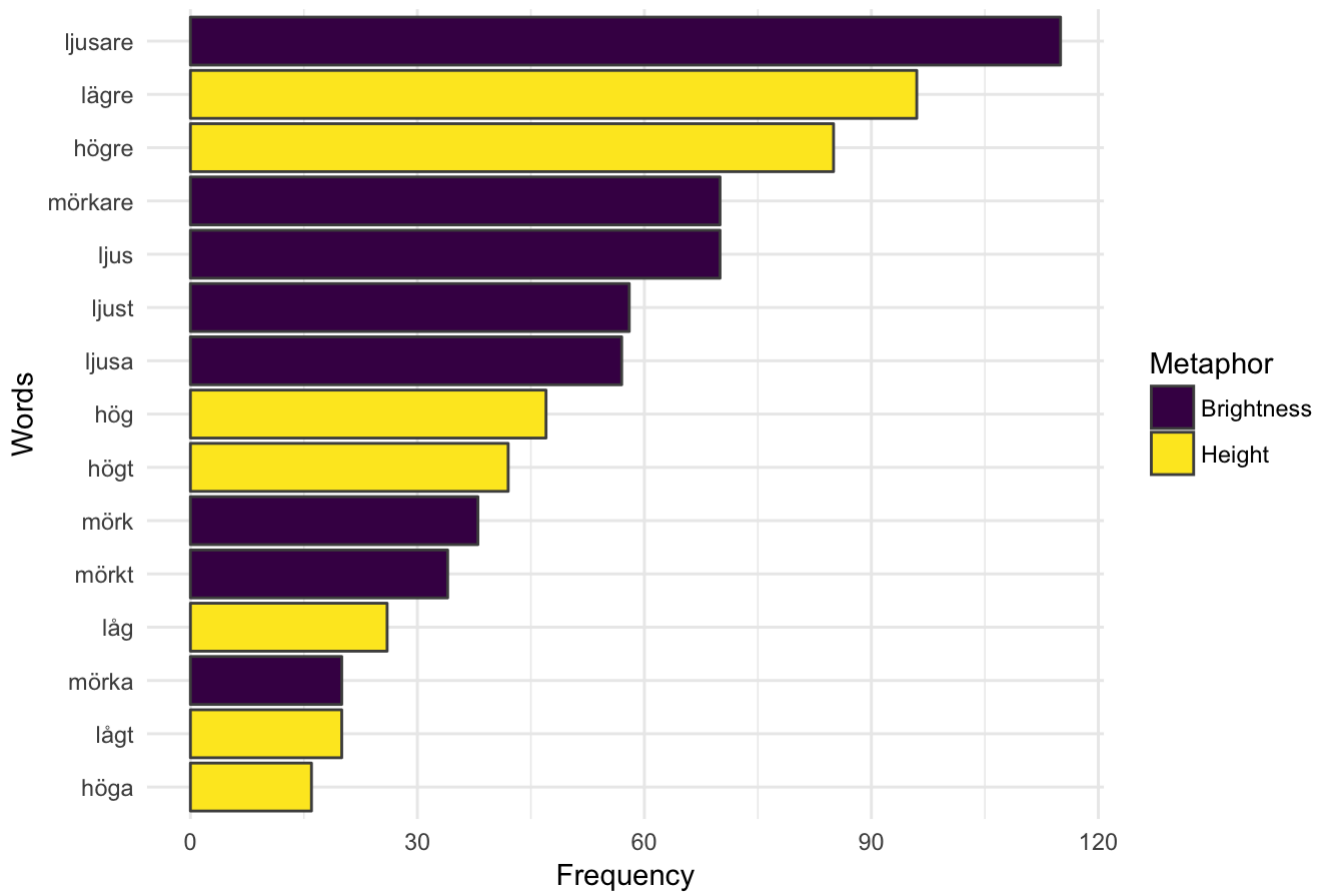
#Swedish
wordsMetaSwe=wordsMeta[wordsMeta$Language=="Swedish"]
wordsMetaSwe[,Freq := .N, by=Words]
wordsMetaSwe=unique(wordsMetaSwe)
wordsMetaSwe=wordsMetaSwe[order(-Freq)]
wordsMetaSwe=wordsMetaSwe[1:15,]

#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

```

Words in Swedish grouped by metaphor



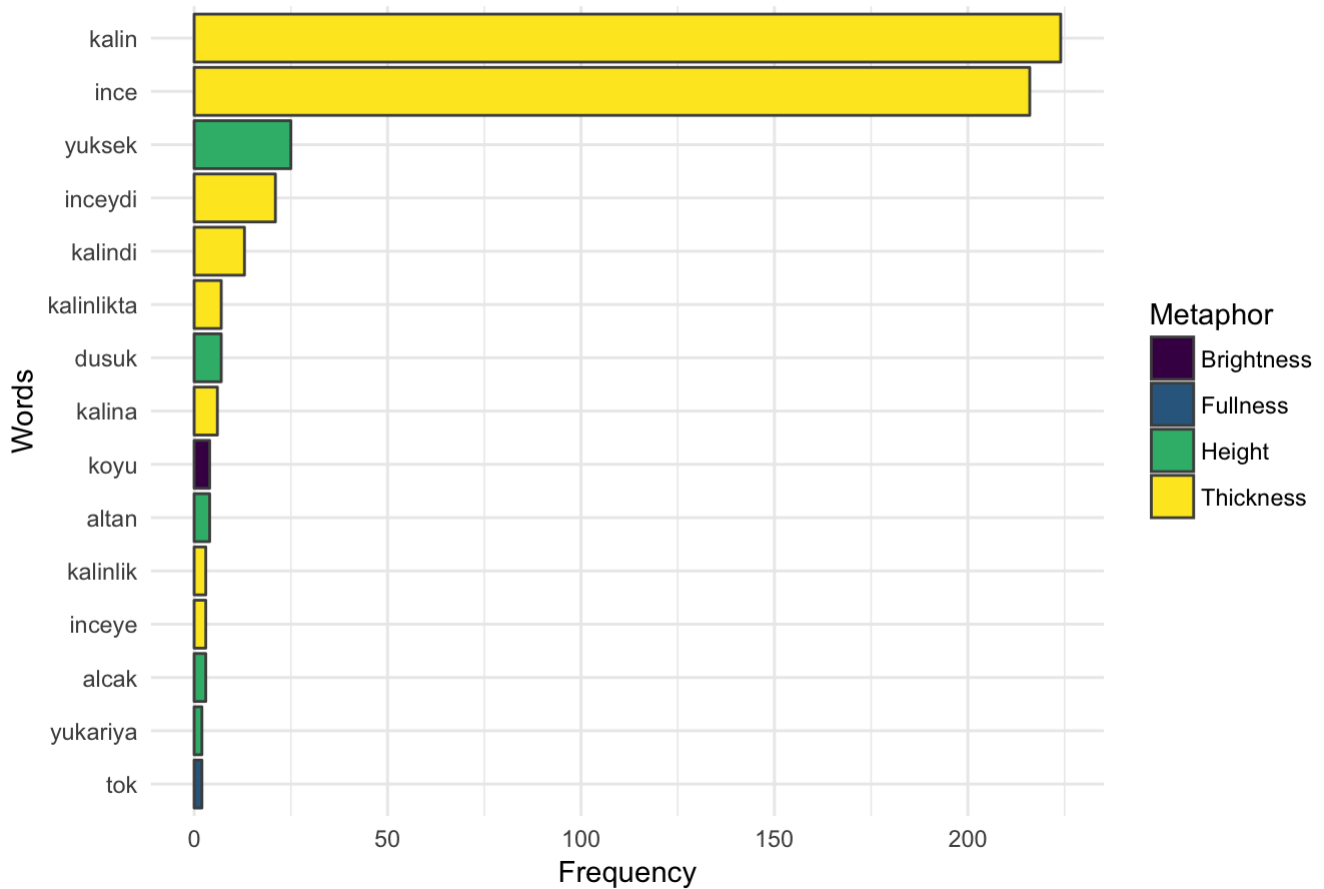
```
#Turkish
```

```
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
```

Words in Turkish grouped by metaphor



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", "yukse
ek"="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"="Ad
j", "yukariya"="Adj-Dat"))
```

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

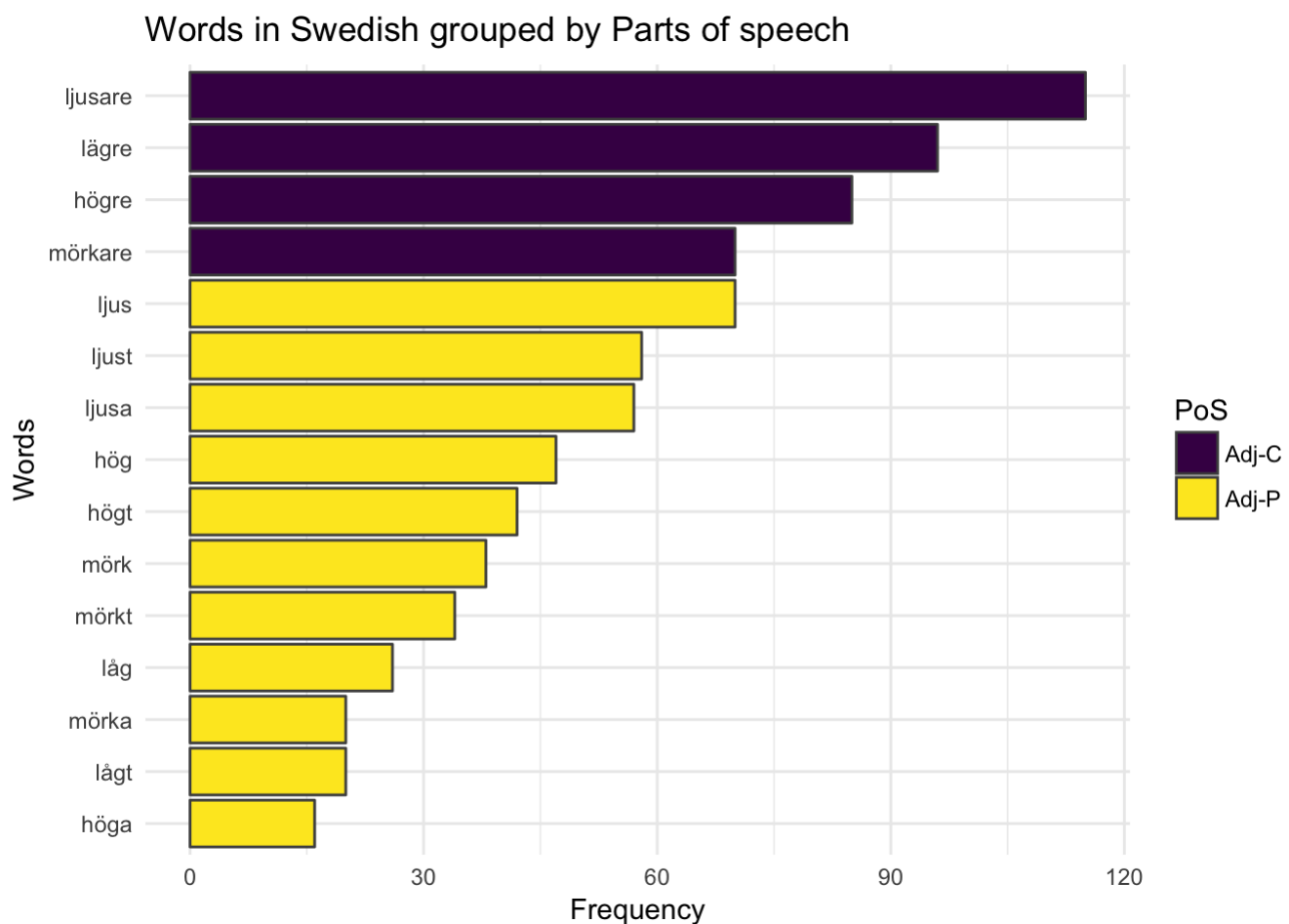
```

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]
```

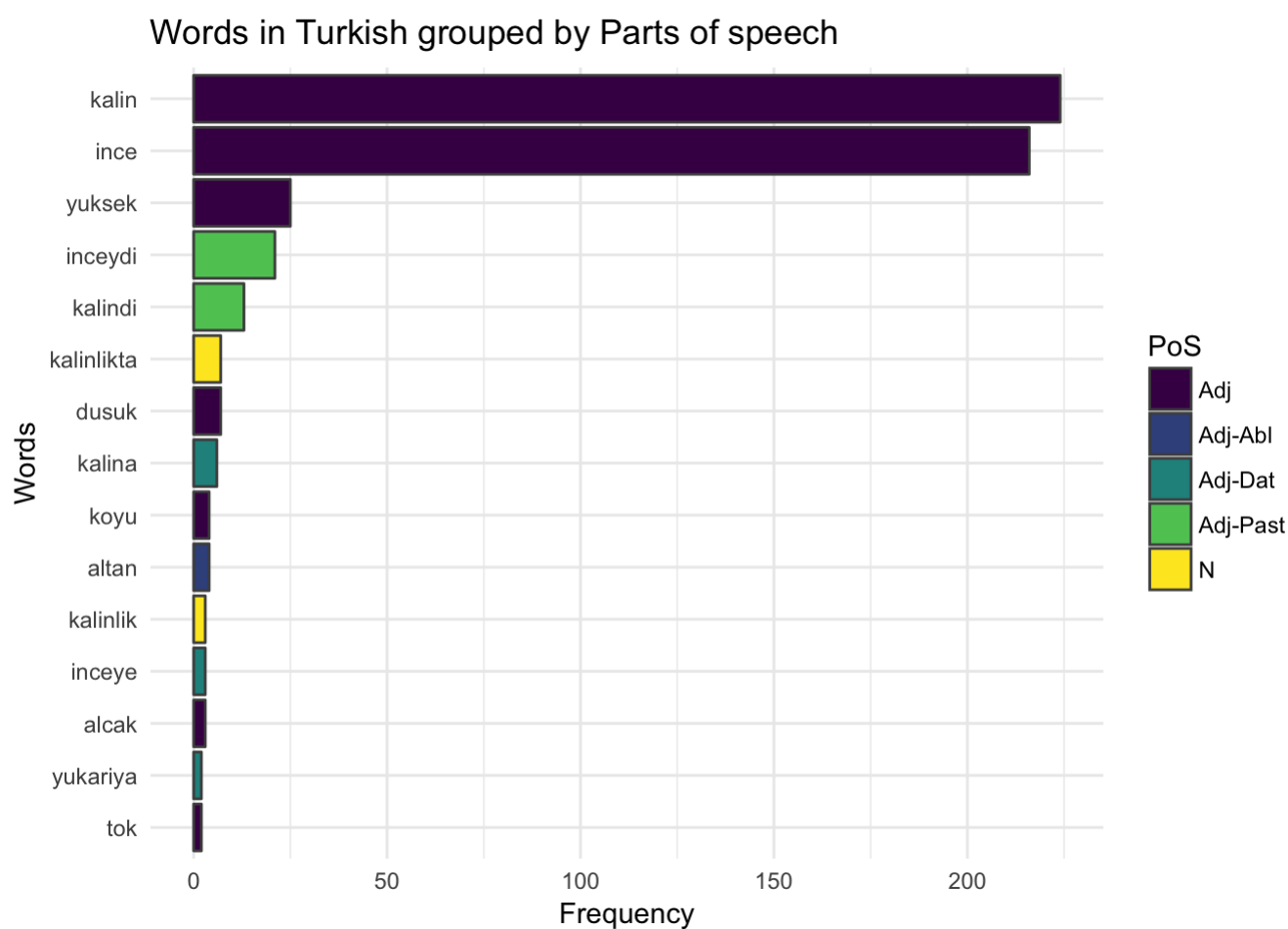
```
wordsPoSTur=unique(wordsPoSTur)
```

```
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"="dee
p",
                                     "högre"="high","mörkare"="dark",
                                     "ljus"="bright","ljus"="bright","ljusa"="brig
ht",
                                     "mörk"="dark","mörka"="dark","mörkt"="dark",
                                     "hög"="high","höjd"="high","högt"="high","hög
a"="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","yukse"="high",
                                     "inceydi"="thin","kalindi"="thick","kalinlikt
a"="thick",

"dusuk"="low","kalina"="thick","koyu"="dark","altan"="below",
                                     "alcak"="low","kalinlik"="thick","inceye"="thi
n","tok"="full","yukariya"="high"))
```

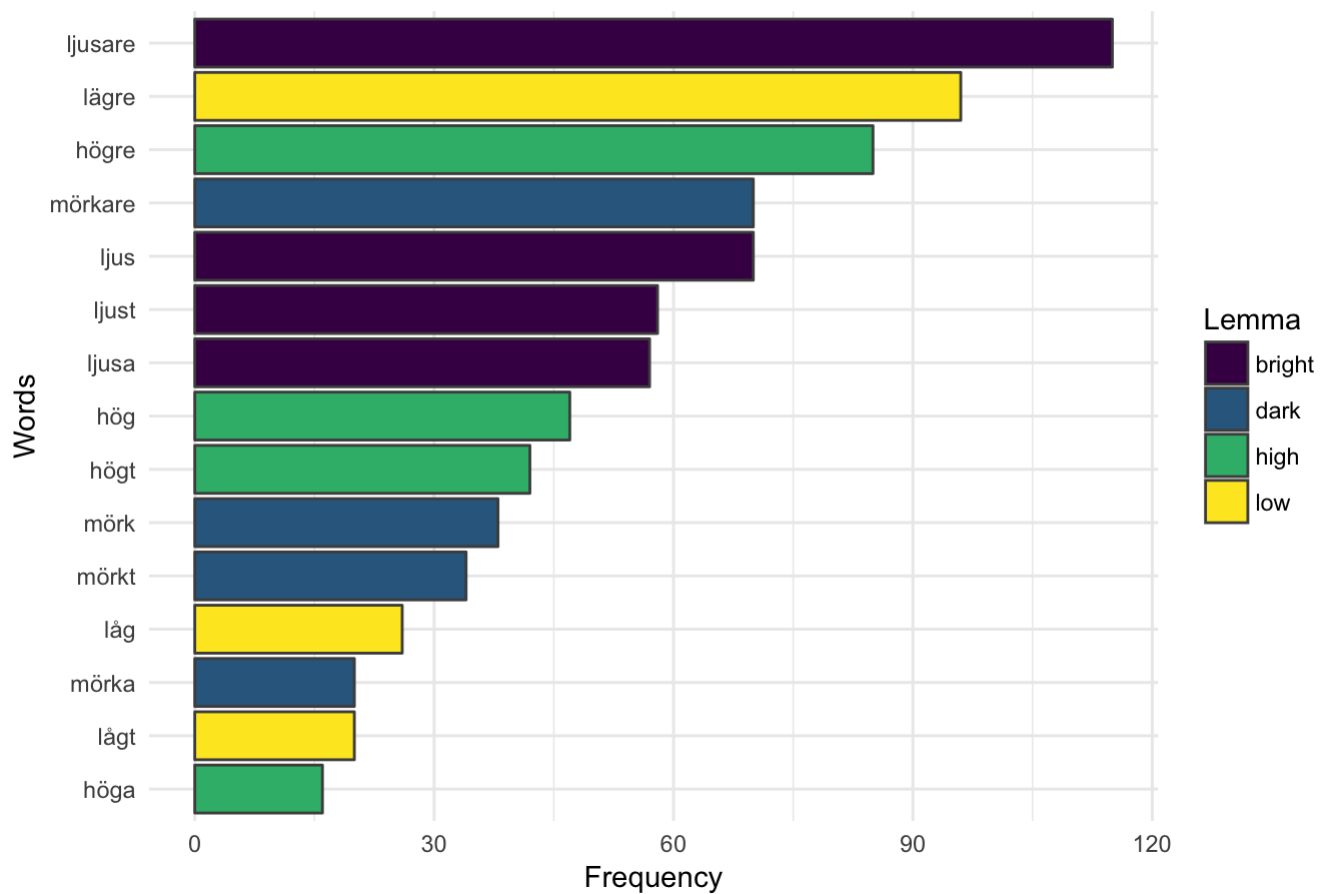
```
## The following `from` values were not present in `x`: över, första
```

```
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
```

Words in Swedish grouped by lemmata



```
#Turkish
```

```
wordsLemTur=wordsLem[wordsLem$Language=="Turkish"]
```

```
wordsLemTur[,Freq := .N, by=Words]
```

```
wordsLemTur=unique(wordsLemTur)
```

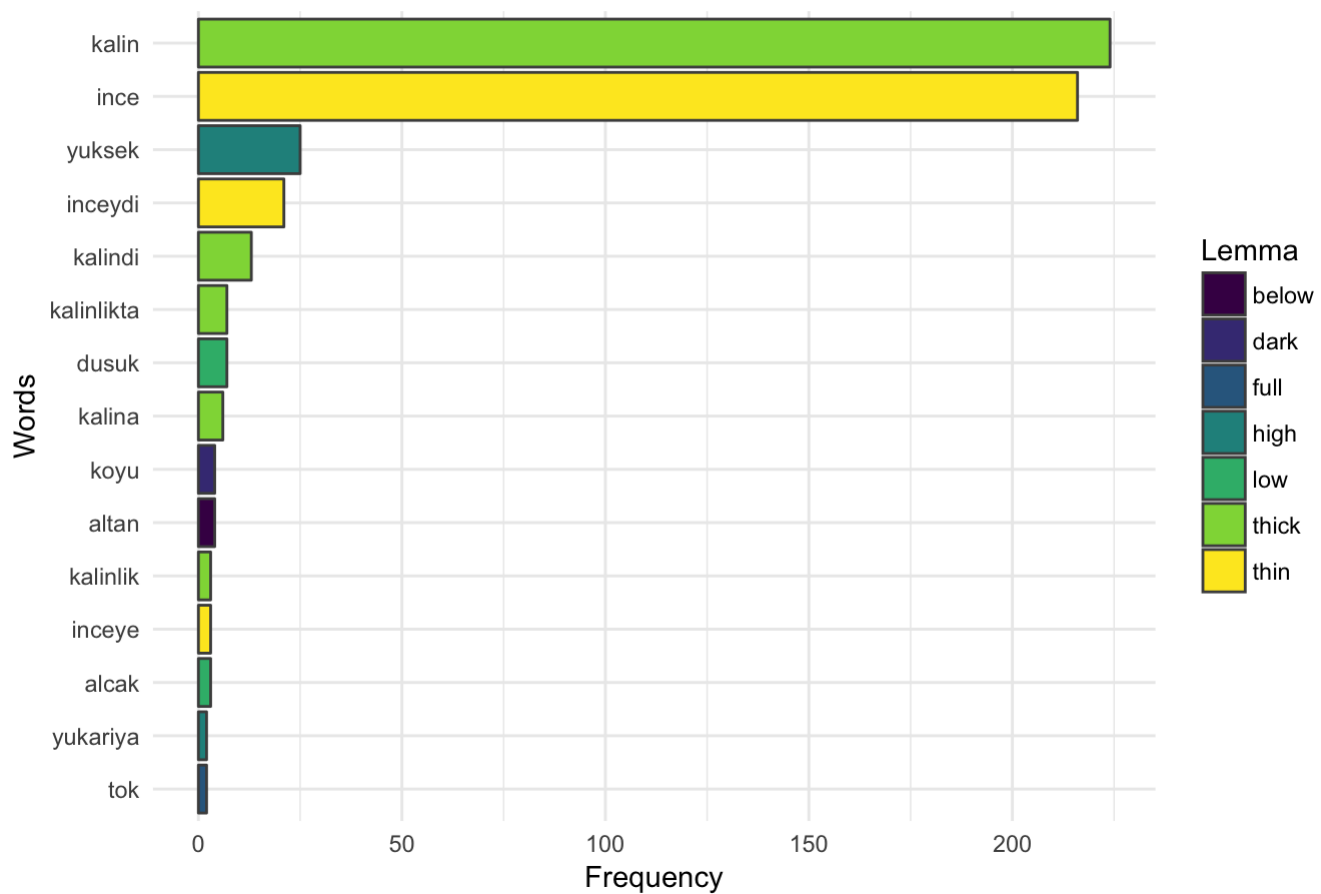
```
wordsLemTur=wordsLemTur[order(-Freq)]
```

```
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
```


Words in Turkish grouped by lemmata

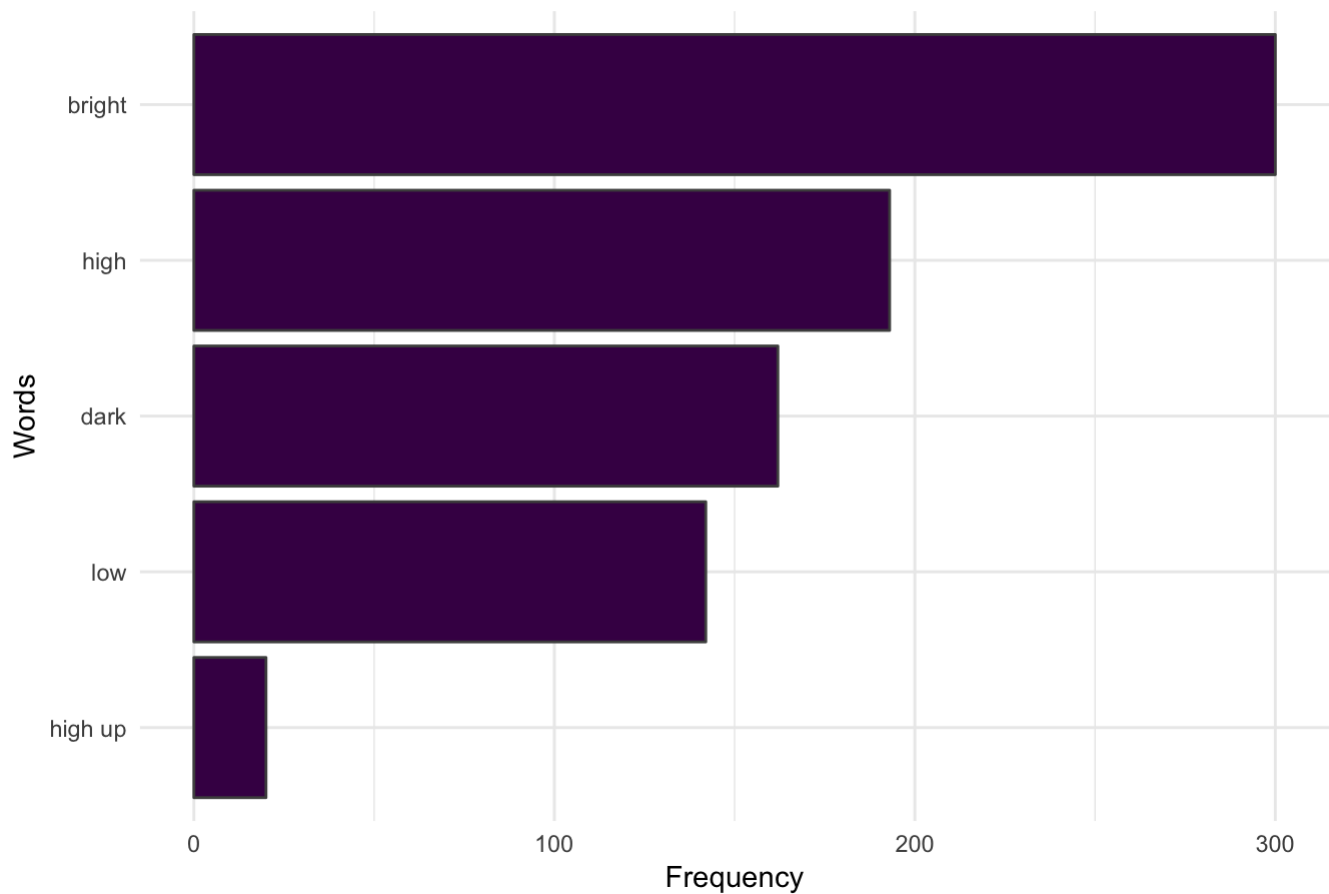


```
### 1.3b Lemmata ###
wordsLem2=wordsLem[, -1]

#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 Turkish") +
  xlab("Words") +
  ylab("Frequency") +
  coord_flip() +
  theme_minimal()
LemSwe2
```

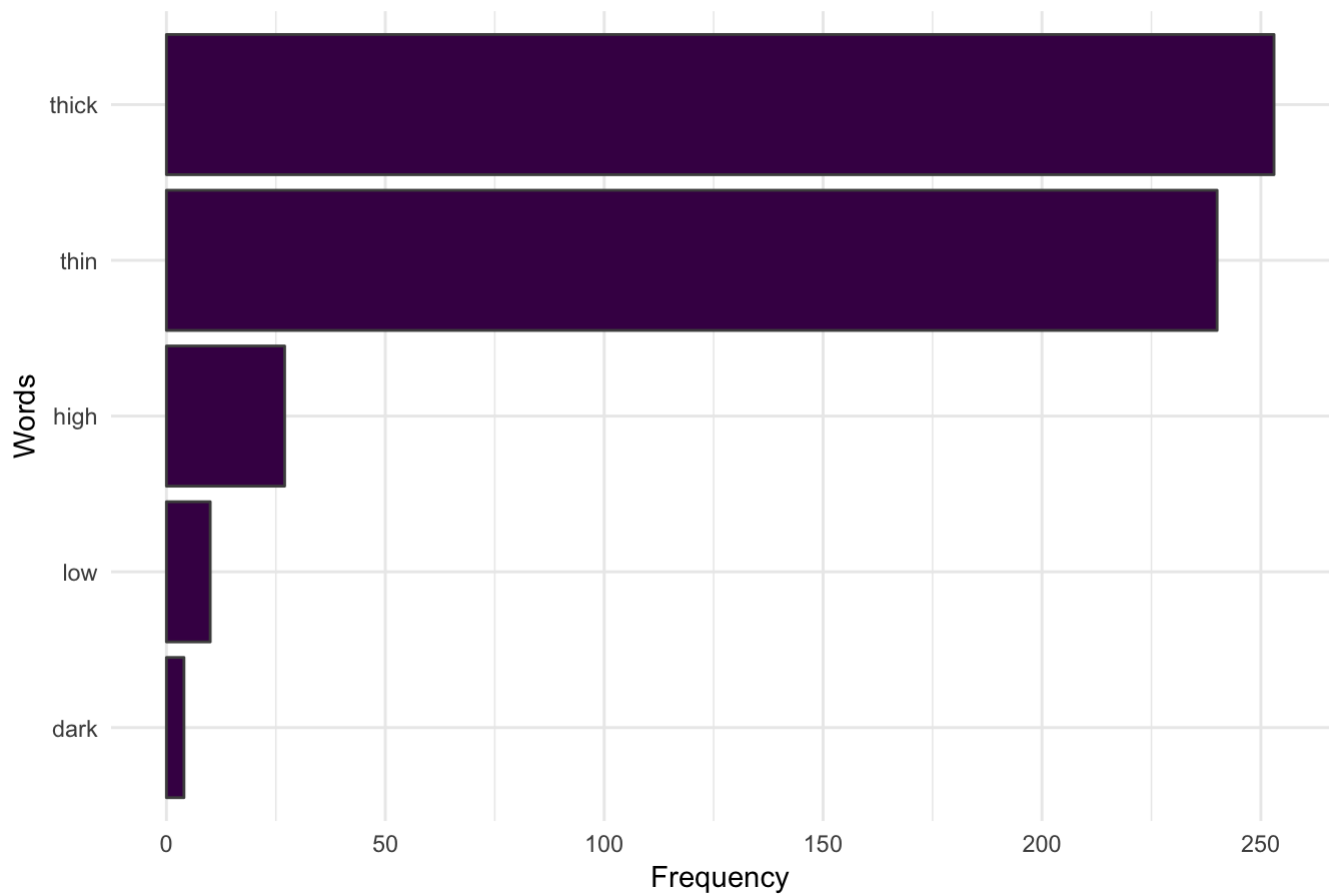
The most frequently used lemmata in Turkish



```
#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
```

The most frequently used lemmata in Turkish



```
##### 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
```

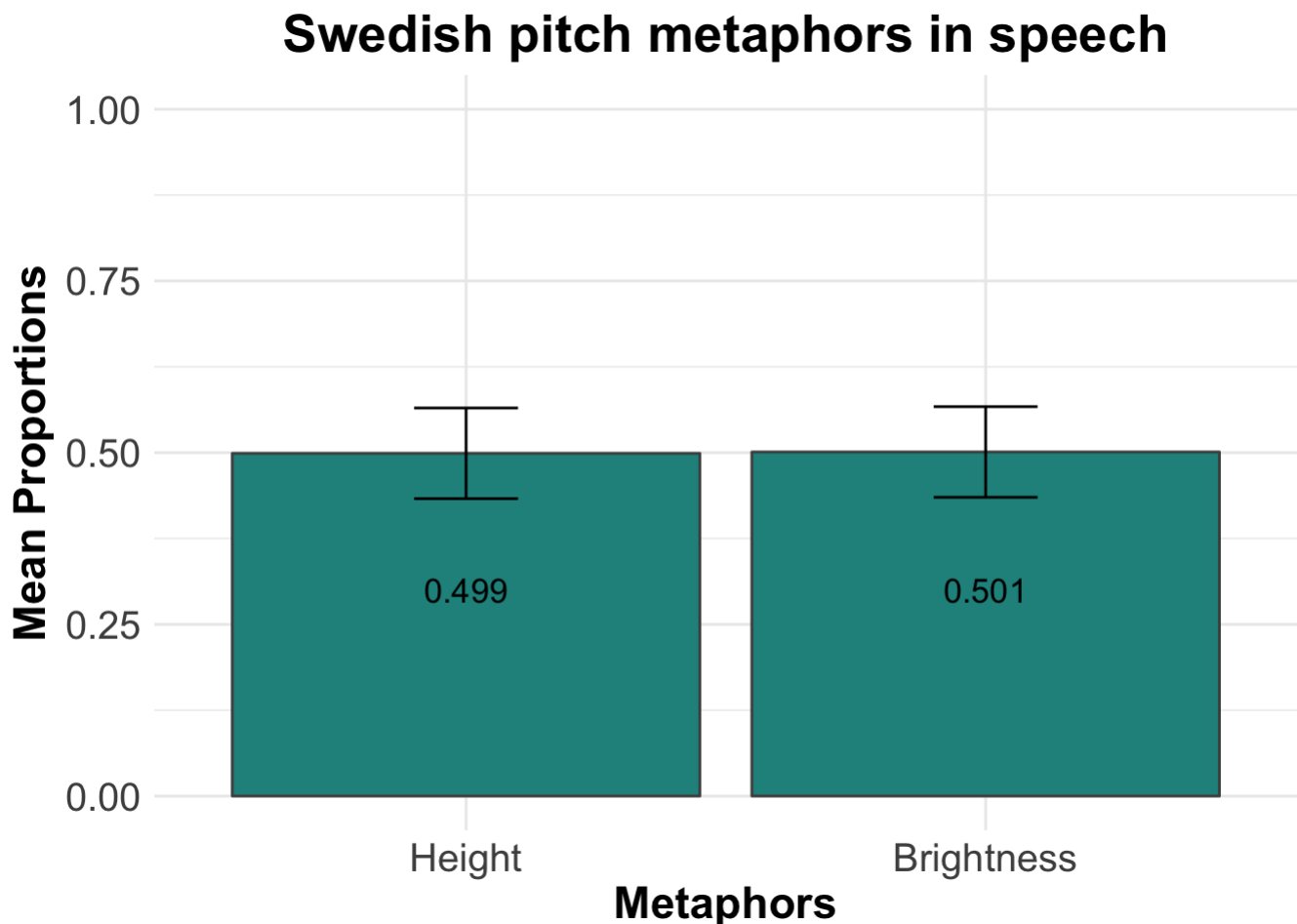
```

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
```

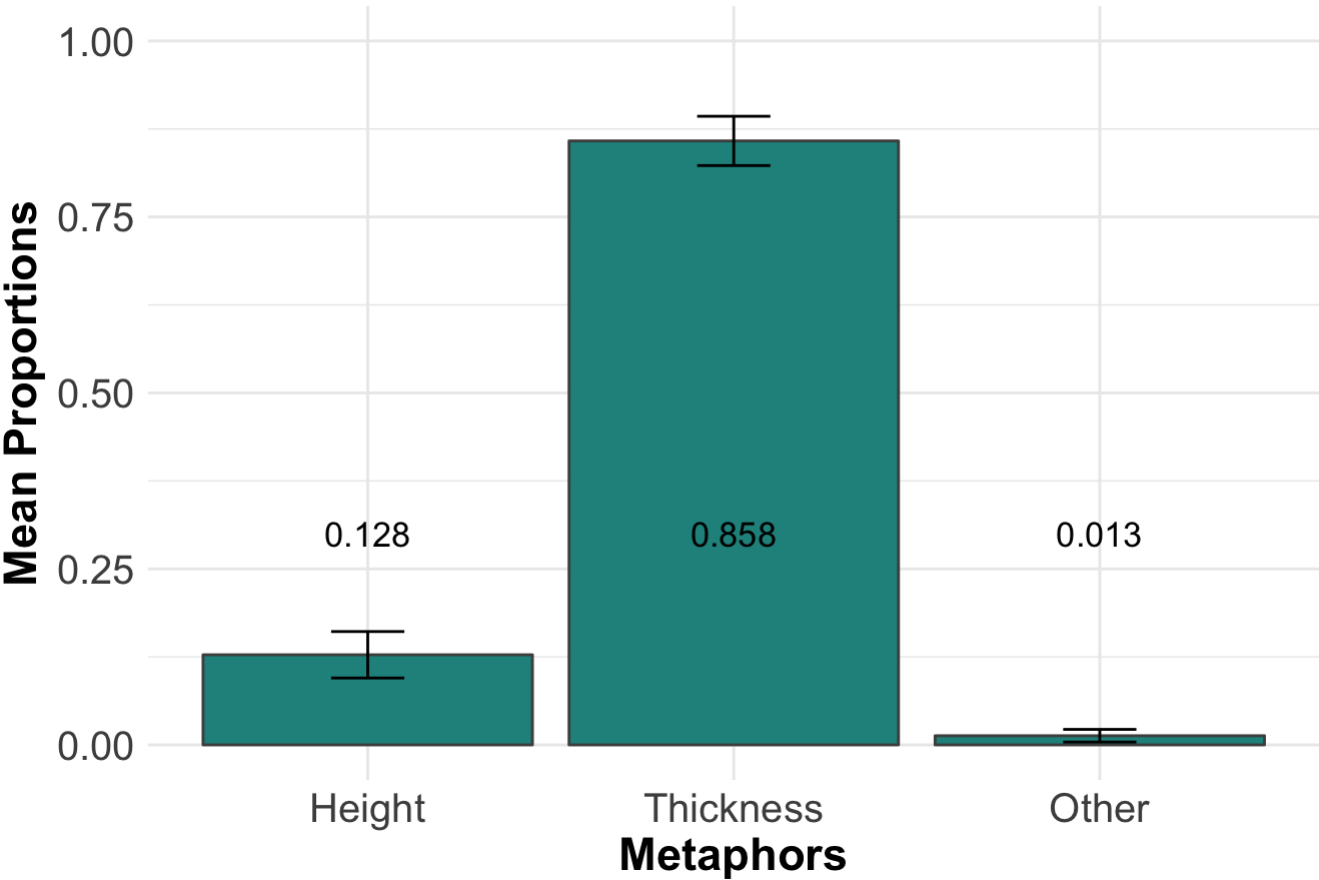
```
dfT=data.frame(cbind(dfT[1:3,],
                     dfT[4:6,2:3]))

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
```

Turkish pitch metaphors in speech



2.2 Consistency over time ####
#Swedish height and Turkish thickness, lineplot with error bars