**ANEXOS**

**Anexo 1 – *Script* para extração e tratamento estatístico dos dados quantitativos no ambiente de programação R**

rm(list=ls(all=T))

*# setwd("/Users/andresouza/Documents/Orientacao/Leo/Arquivos LEO")*

setwd("/Users/andresouza/Documents/Orientacao/Leo/Arquivos LEO")

*# filenames=list.files(path="/Users/andresouza/Documents/Orientacao/Leo/Arquivos LEO/files")*

filenames=list.files(path="C:/Users/fernando/Desktop/Arquivos LEO")

for (a in filenames){

name=gsub(" ","\_",a)

name=gsub(".csv","",name)

*# a=paste("/Users/andresouza/Documents/Orientacao/Leo/Arquivos LEO/files/",a,sep="")*

a=paste("C:/Users/fernando/Desktop/Arquivos LEO/",a,sep="")

assign(name,read.csv2(a, header=FALSE,sep=";"))

}

tudo=rbind(artigo\_academico\_IO,artigo\_academico\_IT,artigo\_academico\_PO,artigo\_academico\_PT,discurso\_politico\_IO,discurso\_politico\_IT,discurso\_politico\_PO,discurso\_politico\_PT,divulgacao\_cientifica\_IO,divulgacao\_cientifica\_IT,divulgacao\_cientifica\_PO,divulgacao\_cientifica\_PT,edu\_IO,edu\_IT,edu\_PO,edu\_PT,ficcao\_IO,ficcao\_IT,ficcao\_PO,ficcao\_PT,man\_instr\_IO,man\_instr\_IT,man\_instr\_PO,man\_instr\_PT,resenha\_IO,resenha\_IT,resenha\_PO,resenha\_PT,tur\_IO,tur\_IT,tur\_PO,tur\_PT)

dados=tudo

dados=na.omit(dados)

colnames(dados)=c("palavras","tags","registro","subcórpora")

dados$registro=gsub("divulgaçao\_cientifica","divulgacao\_cientifica",dados$registro)

dados=dados[dados$tags!="SENT",] *# retira os PONTOS*

dados=dados[dados$tags!="QUOTE",] *#retira os PARENTESES*

dados=dados[dados$tags!="VIRG",] *#retira as VIRGULAS*

dados=dados[dados$tags!="SYM",] *#retira os ASTERISCOS*

dados=dados[dados$tags!="SENT",] *# retira os PONTOS*

dados=dados[dados$tags!="LS",] *# retira os LS*

dados=dados[dados$tags!="POS",] *# retira os POS*

dados=dados[dados$tags!="RP",] *# retira os RP*

dados=dados[dados$tags!="EX",] *# retira os EX*

dados=as.data.frame(lapply(dados,function(x)x[,drop=TRUE])) *## delete unused levels*

tag=gsub("IN ","IN",tag)

tag=gsub("\\bL\\b","ADJ",tag)

dados$tags=tag

dados[238576,2]="ADJ" *## corrige a unica entrada sem tag*

dados$tags=as.fator(dados$tags)

dados$tags=gsub("\\+","\_",dados$tags)

dados$tags=gsub("\\$","S",dados$tags)

dados$tags=gsub("\\bIN/that\\b","preposicao",dados$tags)

*## Correspondência de etiquetas em classes de palavras comuns ##*

dados$tags=gsub("\\bJJ\\b","adjetivo",dados$tags)

dados$tags=gsub("\\bJJR\\b","adjetivo",dados$tags)

dados$tags=gsub("\\bJJS\\b","adjetivo",dados$tags)

dados$tags=gsub("\\bADJ\\b","adjetivo",dados$tags)

dados$tags=gsub("\\bRB\\b","adverbio",dados$tags)

dados$tags=gsub("\\bRBR\\b","adverbio",dados$tags)

dados$tags=gsub("\\bRBS\\b","adverbio",dados$tags) 229

dados$tags=gsub("\\bWRB\\b","adverbio",dados$tags)

dados$tags=gsub("\\bADV\\b","adverbio",dados$tags)

dados$tags=gsub("\\bFW\\b","substantivo",dados$tags)

dados$tags=gsub("\\bNN\\b","substantivo",dados$tags)

dados$tags=gsub("\\bNNS\\b","substantivo",dados$tags)

dados$tags=gsub("\\bNP\\b","substantivo",dados$tags)

dados$tags=gsub("\\bNPS\\b","substantivo",dados$tags)

dados$tags=gsub("\\bF\\b","substantivo",dados$tags)

dados$tags=gsub("\\bB\\b","substantivo",dados$tags)

dados$tags=gsub("\\bNOM\\b","substantivo",dados$tags)

dados$tags=gsub("\\bS\\b","substantivo",dados$tags)

dados$tags=gsub("\\bIN\\b","preposicao",dados$tags)

dados$tags=gsub("\\bTO\\b","preposicao",dados$tags)

dados$tags=gsub("\\bPRP\\b","preposicao",dados$tags)

dados$tags=gsub("\\bPRP\_DET\\b","preposicao",dados$tags)

dados$tags=gsub("\\bPRP\_ADV\\b","preposicao",dados$tags)

dados$tags=gsub("\\bCC\\b","conjuncao",dados$tags)

dados$tags=gsub("\\bCONJ\\b","conjuncao",dados$tags)

dados$tags=gsub("\\bDT\\b","determinante",dados$tags)

dados$tags=gsub("\\bDET\\b","determinante",dados$tags)

dados$tags=gsub("\\bPP\\b","pronome",dados$tags)

dados$tags=gsub("\\bWDT\\b","pronome",dados$tags)

dados$tags=gsub("\\bWP\\b","pronome",dados$tags)

dados$tags=gsub("\\bWPS\\b","pronome",dados$tags)

dados$tags=gsub("\\bPPS\\b","pronome",dados$tags)

dados$tags=gsub("\\bP\\b","pronome",dados$tags)

dados$tags=gsub("\\bPR\\b","pronome",dados$tags)

dados$tags=gsub("\\bCD\\b","numero",dados$tags)

dados$tags=gsub("\\bCARD\\b","numero",dados$tags)

dados$tags=gsub("\\bUH\\b","interjeicao",dados$tags)

dados$tags=gsub("\\bI\\b","interjeicao",dados$tags)

dados$tags=gsub("\\bMD\\b","verbo",dados$tags)

dados$tags=gsub("\\bVB\\b","verbo",dados$tags)

dados$tags=gsub("\\bVBD\\b","verbo",dados$tags)

dados$tags=gsub("\\bVBG\\b","verbo",dados$tags)

dados$tags=gsub("\\bVBN\\b","verbo",dados$tags)

dados$tags=gsub("\\bVBP\\b","verbo",dados$tags)

dados$tags=gsub("\\bVBZ\\b","verbo",dados$tags)

dados$tags=gsub("\\bVVG\\b","verbo",dados$tags)

dados$tags=gsub("\\bVV\\b","verbo",dados$tags)

dados$tags=gsub("\\bVHN\\b","verbo",dados$tags)

dados$tags=gsub("\\bVHZ\\b","verbo",dados$tags)

dados$tags=gsub("\\bVHD\\b","verbo",dados$tags)

dados$tags=gsub("\\bVHG\\b","verbo",dados$tags)

dados$tags=gsub("\\bVVD\\b","verbo",dados$tags)

dados$tags=gsub("\\bVVZ\\b","verbo",dados$tags)

dados$tags=gsub("\\bVH\\b","verbo",dados$tags)

dados$tags=gsub("\\bVVN\\b","verbo",dados$tags)

dados$tags=gsub("\\bVVP\\b","verbo",dados$tags)

dados$tags=gsub("\\bVHP\\b","verbo",dados$tags)

dados$tags=gsub("\\bV\\b","verbo",dados$tags)

dados$tags=gsub("\\bV\_P\\b","verbo",dados$tags)

data.frame(table(dados$tags))

*### Extração de dados##*

*table(dados$tags,dados$subcórpora) ## Essa linha extrai a frequência de palavras por*

*subcórpus*

*table(dados$tags,dados$subcórpora,dados$tags,dados$registro) ## Essa linha extrai a*

*frequência de palavras por registro E subcórpus*

*conj=dados[dados$tags==&quot;conjuncao&quot;,] ## Esta linha reconhece apenas as etiquetas*

*referentes às conjunções*

*conj=as.data.frame(lapply(conj,function(x)x[,drop=TRUE])) ## delete unused levels*

*table(conj$registro,conj$subcórpora) # Esta linha extrai a frequência de conjunções por*

*registro E subcórpus*

*##Testes de significância##*

*x=c(....) # coloque os números aqui por linha*

*tabela=matrix(x,2,8,byrow=T)*

*chisq.test(tabela) # Teste qui-quadrado por aderência*

*chisq.test(tabela)$stdres # Teste post hoc Z*