

qorder

Michael Pearson

2/23/2018

Remove the one-offs

now let's process the ones with multiple bigrams

```
blocky <- function(trap, tim, ful_tri) {  
  a <- floor(nrow(tim)/1000)  
  full_tri <- data.table()  
  downstream <- 0.5  
  for (i in 1:a)  
  {  
    tardis <- combi_tri_ns_ns[combi_tri_ns_ns$bigrams == c(tim[i]),]  
    tardis$sum <- sum(tardis$tri_gram_ns_ns)  
    tardis$prob <- ((tardis$tri_gram_ns_ns - downstream)/tardis$sum)  
    full_tri <- rbind(full_tri, tardis)  
  }  
  return(full_tri)  
}  
  
combi_tri_ns_ns <- read.csv("/Users/mutecypher/Documents/Documents - Michael's iMac/Coursera/Capstone P  
combi_tri_ns_ns <- data.table(combi_tri_ns_ns)  
trixy <- combi_tri_ns_ns[combi_tri_ns_ns$tri_gram_ns_ns >= 2,]  
tim <- unique(trixy$bigrams)  
tim <- as.character(tim)  
tim <- data.table(tim)  
blah <- blocky(trixy, tim, full_tri)  
write.csv(blah, file = "/Users/mutecypher/Documents/Documents - Michael's iMac/Coursera/Capstone Project,
```

quadgrams