X-much in IrE and AusE - Part 5: Statistical Analysis

Anonymous

2022-01-02

This document focuses on the use of the X-much structure in Irish and Australian English.

This part of the analysis performs the data analysis. The data is processed using *tidyverse* package(s) (Wickham et al. [2019](#ref-tidyverse)) and the statistical analysis uses 2-tests (see Cochran [1952](#ref-cochran1952x2)) and configural frequency analysis (see Lienert and Krauth [1975](#ref-lienert1975cfa)) using the *cfa* package (Mair and Funke [2017](#ref-cfa)). Effect sizes are calculated using the *effectsize* package (Ben-Shachar, Lüdecke, and Makowski [2020](#ref-effectsize)).

# Session preparation

* install packages

install.packages("tidyverse")  
install.packages("here")  
install.packages("readxl")  
install.packages("flextable")  
install.packages("quanteda")  
install.packages("tidytext")  
install.packages("cfa")  
install.packages("report")  
install.packages("effectsize")

* load packages
* set options

# load packages  
library(tidyverse)

## -- Attaching packages --------------------------------------- tidyverse 1.3.1 --

## v ggplot2 3.3.5 v purrr 0.3.4  
## v tibble 3.1.6 v dplyr 1.0.7  
## v tidyr 1.1.4 v stringr 1.4.0  
## v readr 2.1.1 v forcats 0.5.1

## Warning: Paket 'tibble' wurde unter R Version 4.1.2 erstellt

## Warning: Paket 'readr' wurde unter R Version 4.1.2 erstellt

## -- Conflicts ------------------------------------------ tidyverse\_conflicts() --  
## x dplyr::filter() masks stats::filter()  
## x dplyr::lag() masks stats::lag()

library(here)

## here() starts at D:/Uni/Projekte/XmuchIrE

library(readxl)  
library(flextable)

## Warning: Paket 'flextable' wurde unter R Version 4.1.2 erstellt

##   
## Attache Paket: 'flextable'

## Das folgende Objekt ist maskiert 'package:purrr':  
##   
## compose

library(quanteda)

## Warning: Paket 'quanteda' wurde unter R Version 4.1.2 erstellt

## Package version: 3.2.0  
## Unicode version: 13.0  
## ICU version: 69.1

## Parallel computing: 8 of 8 threads used.

## See https://quanteda.io for tutorials and examples.

library(tidytext)  
library(cfa)  
library(vcd)

## Lade nötiges Paket: grid

library(effectsize)

## Warning: Paket 'effectsize' wurde unter R Version 4.1.2 erstellt

##   
## Attache Paket: 'effectsize'

## Das folgende Objekt ist maskiert 'package:vcd':  
##   
## oddsratio

# setting options  
options(stringsAsFactors = F)

# Load data

xmuch <- base::readRDS(file = here::here("data", "xmuch.rda"))  
# inspect  
head(xmuch, 10)

## docname Variety  
## 1 wlp\_au\_b01\_##3282941.1 Australia  
## 2 wlp\_au\_b01\_##3287006.1 Australia  
## 3 wlp\_au\_b01\_##3289741.1 Australia  
## 4 wlp\_au\_b01\_##3291338.1 Australia  
## 5 wlp\_au\_b01\_##3293441.1 Australia  
## 6 wlp\_au\_b01\_##3297127.1 Australia  
## 7 wlp\_au\_b01\_##3304019.1 Australia  
## 8 wlp\_au\_b01\_##3304123.1 Australia  
## 9 wlp\_au\_b01\_##3304829.1 Australia  
## 10 wlp\_au\_b01\_##3306913.1 Australia  
## kwic  
## 1 lessons . Evi 's first day at home and her first collar ( 8 weeks old ) ./. Cute << much >> ? Evi first seizure . In comparison to what  
## 2 a ring fire with cotton balls and made ' marshmallows on sticks ' cotton balls on pencil tips . Creative << much >> ? ?/? So , we sat fireside &; roasted marshmallows &; all crammed into a play tent to sleep &;  
## 3 the skies will fall if we have a carbon tax . " Skies will fall " ?/? Hype , << much >> ? I do n't think anyone here asserted that . It 's a supid policy that will distort the economy  
## 4 the sleek uniform of the bowl cut which was styled for Lumley by John Frida in 1976 . Twins << much >> ? With the hair cut I became fixed on the clothes  
## 5 powder served alongside it too much lately , I even sneak spoonfuls out of the fridge - pregnancy craving << much >> ? It 's just 1 macadamia nuts shaved on the micro-plane , 1 cup of nutritional yeast and 1 tsp  
## 6 the world a round of psychotherapy to open their eyes to the truth of the world ( um .../... feasible << much >> ? !/! ) ,/, I do n't see any other solution . As someone who has suffered from depression for  
## 7 teach . Can you listen ? Can you hear them ? Airy fairy << much >> ? .../... here 's the biggest joke . The very people who claim they want connection with the  
## 8 Striesand Effect . Do they think anyone would have seen this piddly little video absent their protests ? Ironic , << much >> ? LE 6- A reasonable point of view . But what should we do from a public policy perspective  
## 9 was so lovely to see these girls , hear all about Sharon 's trip to France and Italy ( jealous << much >> ? )/) and talk just a little bit about photography . Must get some practice in before our next catchup  
## 10 Tweeting Without Getting Sued , which I 'll be buying as soon as it comes out next month ( obsessed << much >> ? )/) . If you have any questions , I can try to answer them or at least point  
## id X POS Status Emotionality Data Polarity  
## 1 1 cute Adjective word emotional Xmuch positive  
## 2 5 creative Adjective word emotional Xmuch positive  
## 3 7 hype Noun word nonemotional Xmuch neutral  
## 4 8 twins Noun word nonemotional Xmuch neutral  
## 5 11 pregnancy craving Noun phrase nonemotional Xmuch neutral  
## 6 17 feasible Adjective word nonemotional Xmuch neutral  
## 7 26 airy fairy Noun phrase nonemotional Xmuch neutral  
## 8 27 ironic Adjective word nonemotional Xmuch neutral  
## 9 29 jealous Adjective word emotional Xmuch negative  
## 10 32 obsessed Adjective word emotional Xmuch negative

Inspect structure of the data

str(xmuch)

## 'data.frame': 2142 obs. of 10 variables:  
## $ docname : chr "wlp\_au\_b01\_##3282941.1" "wlp\_au\_b01\_##3287006.1" "wlp\_au\_b01\_##3289741.1" "wlp\_au\_b01\_##3291338.1" ...  
## $ Variety : chr "Australia" "Australia" "Australia" "Australia" ...  
## $ kwic : chr "lessons . Evi 's first day at home and her first collar ( 8 weeks old ) ./. Cute << much >> ? Evi first seizure"| \_\_truncated\_\_ "a ring fire with cotton balls and made ' marshmallows on sticks ' cotton balls on pencil tips . Creative << muc"| \_\_truncated\_\_ "the skies will fall if we have a carbon tax . \" Skies will fall \" ?/? Hype , << much >> ? I do n't think anyo"| \_\_truncated\_\_ "the sleek uniform of the bowl cut which was styled for Lumley by John Frida in 1976 . Twins << much >> ? With t"| \_\_truncated\_\_ ...  
## $ id : int 1 5 7 8 11 17 26 27 29 32 ...  
## $ X : chr "cute" "creative" "hype" "twins" ...  
## $ POS : chr "Adjective" "Adjective" "Noun" "Noun" ...  
## $ Status : chr "word" "word" "word" "word" ...  
## $ Emotionality: chr "emotional" "emotional" "nonemotional" "nonemotional" ...  
## $ Data : chr "Xmuch" "Xmuch" "Xmuch" "Xmuch" ...  
## $ Polarity : chr "positive" "positive" "neutral" "neutral" ...

# Statistical Analysis

## Single vs Multiword

RQ: do ire and aus differ regarding the status of x (single word vs multiword)?

phrasetb <- xmuch %>%  
 dplyr::filter(Data == "Xmuch") %>%  
 dplyr::select(Variety, Status) %>%  
 dplyr::group\_by(Variety, Status) %>%  
 dplyr::summarise(Frequency = n()) %>%  
 dplyr::group\_by(Variety) %>%  
 dplyr::mutate(Total = sum(Frequency)) %>%  
 dplyr::ungroup() %>%  
 dplyr::rowwise() %>%  
 dplyr::mutate(Percent = round(Frequency/Total\*100, 2))

## `summarise()` has grouped output by 'Variety'. You can override using the `.groups` argument.

# inspect  
phrasetb

## # A tibble: 4 x 5  
## # Rowwise:   
## Variety Status Frequency Total Percent  
## <chr> <chr> <int> <int> <dbl>  
## 1 Australia phrase 31 109 28.4  
## 2 Australia word 78 109 71.6  
## 3 Ireland phrase 10 33 30.3  
## 4 Ireland word 23 33 69.7

phrase\_x2 <- phrasetb %>%  
 dplyr::select(-Total, -Percent) %>%  
 tidyr::spread(Status, Frequency) %>%  
 as.matrix()  
# add rownames  
rownames(phrase\_x2) <- phrase\_x2[, 1]  
phrase\_x2 <- phrase\_x2[, 2:3]  
# convert to numeric  
phrase\_x2 <- t(apply(phrase\_x2, 1, function(x){   
 x <- as.numeric(x) }))  
# add column names  
#colnames(phrase\_x2) <- names(table(phrasetb$element))  
colnames(phrase\_x2) <- c("single-word", "multi-word exp.")  
# inspect  
phrase\_x2

## single-word multi-word exp.  
## Australia 31 78  
## Ireland 10 23

* perform X2-test (R Core Team [2021](#ref-stats))

chisq.test(phrase\_x2)

##   
## Pearson's Chi-squared test with Yates' continuity correction  
##   
## data: phrase\_x2  
## X-squared = 2.2635e-30, df = 1, p-value = 1

# effect size  
effectsize::effectsize(chisq.test(phrase\_x2), type = "phi")

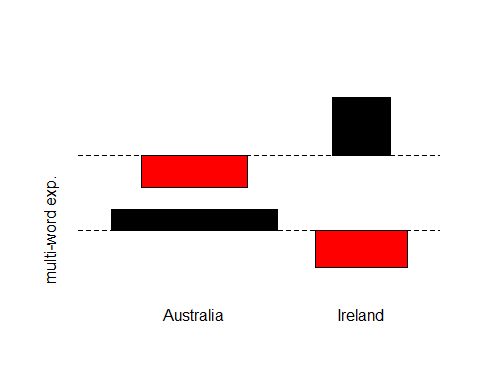
## Phi | 95% CI  
## -------------------  
## 0.02 | [0.00, 1.00]  
##   
## - One-sided CIs: upper bound fixed at (1).

generate visualization

# open window  
png(here::here("images", "multiword\_assoc.png"), width = 500, height = 450)  
# generate plot  
assocplot(phrase\_x2)  
# close window  
dev.off()

## png   
## 2

assocplot(phrase\_x2)



## Emotionality

tabulation

emo\_tab <- xmuch %>%  
 dplyr::group\_by(Variety, Status, Polarity) %>%  
 dplyr::summarise(Freq = n()) %>%  
 dplyr::group\_by(Variety, Status) %>%  
 dplyr::mutate(Total = sum(Freq)) %>%  
 dplyr::rowwise() %>%  
 dplyr::mutate(Percent = round(Freq/Total\*100, 1),  
 Frequency = paste0(Freq, " (", Percent, ")")) %>%  
 dplyr::ungroup() %>%  
 dplyr::select(-Freq, -Total, -Percent) %>%  
 tidyr::spread(Polarity, Frequency)

## `summarise()` has grouped output by 'Variety', 'Status'. You can override using the `.groups` argument.

# inspect  
emo\_tab

## # A tibble: 4 x 5  
## Variety Status negative neutral positive  
## <chr> <chr> <chr> <chr> <chr>   
## 1 Australia phrase 11 (35.5) 20 (64.5) <NA>   
## 2 Australia word 66 (6.1) 933 (86.5) 79 (7.3)  
## 3 Ireland phrase 4 (40) 6 (60) <NA>   
## 4 Ireland word 36 (3.5) 939 (91.8) 48 (4.7)

Calculate totals

# au  
austot <- 11+63+20+952+0+63  
negau <- round(74/austot\*100, 1)  
nonau <- round(972/austot\*100, 1)  
posau <- round(63/austot\*100, 1)  
# ire  
iretot <- 4+37+6+929+0+57  
negire <- round(41/iretot\*100, 1)  
nonire <- round(935/iretot\*100, 1)  
posire <- round(57/iretot\*100, 1)  
# results  
negau

## [1] 6.7

nonau

## [1] 87.6

posau

## [1] 5.7

negire

## [1] 4

nonire

## [1] 90.5

posire

## [1] 5.5

xmdtb <- xmuch %>%  
 dplyr::select(Variety, X, POS, Emotionality, Polarity, Status, Data) %>%  
 dplyr::rename(word = X)  
# inspect  
head(xmdtb)

## Variety word POS Emotionality Polarity Status Data  
## 1 Australia cute Adjective emotional positive word Xmuch  
## 2 Australia creative Adjective emotional positive word Xmuch  
## 3 Australia hype Noun nonemotional neutral word Xmuch  
## 4 Australia twins Noun nonemotional neutral word Xmuch  
## 5 Australia pregnancy craving Noun nonemotional neutral phrase Xmuch  
## 6 Australia feasible Adjective nonemotional neutral word Xmuch

* tabulation

tb1 <-base::readRDS(file = here::here("tables", "tb1.rda")) %>%  
 dplyr::mutate(Variety = stringr::str\_replace\_all(text, ".\*\_([a-z]{2,2})\_.\*", "\\1"),  
 Variety = ifelse(Variety == "au", "Australia", "Ireland")) %>%  
 dplyr::group\_by(Variety) %>%  
 dplyr::summarise(words = sum(words),  
 texts = n())  
# inspect  
head(tb1)

## # A tibble: 2 x 3  
## Variety words texts  
## <chr> <int> <int>  
## 1 Australia 152077983 129382  
## 2 Ireland 103259885 102426

xmdtb %>%  
 dplyr::group\_by(Variety, Status) %>%  
 dplyr::summarise(Frequency = n())

## `summarise()` has grouped output by 'Variety'. You can override using the `.groups` argument.

## # A tibble: 4 x 3  
## # Groups: Variety [2]  
## Variety Status Frequency  
## <chr> <chr> <int>  
## 1 Australia phrase 31  
## 2 Australia word 1078  
## 3 Ireland phrase 10  
## 4 Ireland word 1023

### X2 (wo variety)

RQ: Does the emotionality of words in the test data (in the x-much construction) differ from the emotionality in the control data?

emo\_tb1 <- xmdtb %>%  
 dplyr::group\_by(Emotionality, Data) %>%  
 dplyr::summarise(Frequency = n())

## `summarise()` has grouped output by 'Emotionality'. You can override using the `.groups` argument.

# inspect  
head(emo\_tb1)

## # A tibble: 4 x 3  
## # Groups: Emotionality [2]  
## Emotionality Data Frequency  
## <chr> <chr> <int>  
## 1 emotional control 169  
## 2 emotional Xmuch 75  
## 3 nonemotional control 1831  
## 4 nonemotional Xmuch 67

Perform x2-test

emo\_x2 <- emo\_tb1 %>%  
 tidyr::spread(Data, Frequency) %>%  
 as.matrix()  
rn<- emo\_x2[,1]  
emo\_x2 = emo\_x2[, 2:3]  
emo\_x2 <- apply(emo\_x2, 2, as.numeric)  
rownames(emo\_x2) <- rn  
emo\_x2

## control Xmuch  
## emotional 169 75  
## nonemotional 1831 67

# perform x2 test  
chisq.test(emo\_x2)

##   
## Pearson's Chi-squared test with Yates' continuity correction  
##   
## data: emo\_x2  
## X-squared = 254.19, df = 1, p-value < 2.2e-16

# effect size  
effectsize::effectsize(chisq.test(emo\_x2), type = "phi")

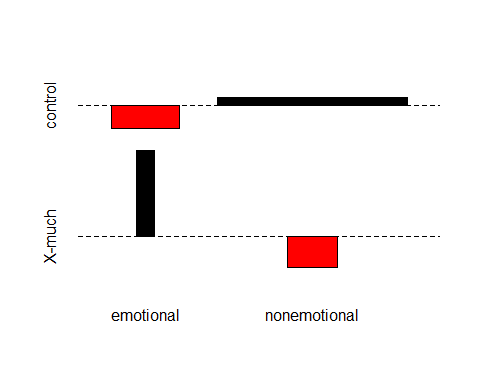
## Phi | 95% CI  
## -------------------  
## 0.35 | [0.31, 1.00]  
##   
## - One-sided CIs: upper bound fixed at (1).

The words in the x-much structure are significantly more emotional compared to the words in the control data.

# adapt column names  
colnames(emo\_x2) <- c("control", "X-much")  
  
# open window  
png(here::here("images", "emo\_assoc.png"), width = 500, height = 350)  
# generate plot  
assocplot(emo\_x2)  
# close window  
dev.off()

## png   
## 2

assocplot(emo\_x2)



### X2 (vareity)

RQ: Does the emotionality of words in the X-much structure in Australia differ from the emotionality of words in the X-much structure in Ireland?

emo\_tb2 <- xmuch %>%  
 dplyr::filter(Data == "Xmuch") %>%  
 dplyr::group\_by(Emotionality, Variety) %>%  
 dplyr::summarise(Frequency = n())

## `summarise()` has grouped output by 'Emotionality'. You can override using the `.groups` argument.

# inspect  
head(emo\_tb2)

## # A tibble: 4 x 3  
## # Groups: Emotionality [2]  
## Emotionality Variety Frequency  
## <chr> <chr> <int>  
## 1 emotional Australia 60  
## 2 emotional Ireland 15  
## 3 nonemotional Australia 49  
## 4 nonemotional Ireland 18

Perform x2-test

emo\_x22 <- emo\_tb2 %>%  
 tidyr::spread(Variety, Frequency) %>%  
 as.matrix()  
rn<- emo\_x22[,1]  
emo\_x22 = emo\_x22[, 2:3]  
emo\_x22 <- apply(emo\_x22, 2, as.numeric)  
rownames(emo\_x22) <- rn  
emo\_x22

## Australia Ireland  
## emotional 60 15  
## nonemotional 49 18

# perform x2 test  
chisq.test(emo\_x22)

##   
## Pearson's Chi-squared test with Yates' continuity correction  
##   
## data: emo\_x22  
## X-squared = 0.58981, df = 1, p-value = 0.4425

# effect size  
effectsize::effectsize(chisq.test(emo\_x22), type = "phi")

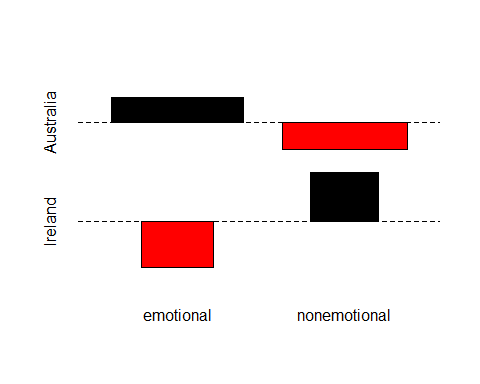
## Phi | 95% CI  
## -------------------  
## 0.08 | [0.00, 1.00]  
##   
## - One-sided CIs: upper bound fixed at (1).

Australia and Ireland differ significantly with respect to the emotionality of words in the x-much structure.

# open window  
png(here::here("images", "emo\_var\_assoc.png"), width = 500, height = 350)  
# generate plot  
assocplot(emo\_x22)  
# close window  
dev.off()

## png   
## 2

assocplot(emo\_x22)



### CFA

emo\_tb3 <- xmdtb %>%  
 dplyr::group\_by(Emotionality, Variety, Data) %>%  
 dplyr::summarise(Frequency = n())

## `summarise()` has grouped output by 'Emotionality', 'Variety'. You can override using the `.groups` argument.

# inspect  
emo\_tb3

## # A tibble: 8 x 4  
## # Groups: Emotionality, Variety [4]  
## Emotionality Variety Data Frequency  
## <chr> <chr> <chr> <int>  
## 1 emotional Australia control 96  
## 2 emotional Australia Xmuch 60  
## 3 emotional Ireland control 73  
## 4 emotional Ireland Xmuch 15  
## 5 nonemotional Australia control 904  
## 6 nonemotional Australia Xmuch 49  
## 7 nonemotional Ireland control 927  
## 8 nonemotional Ireland Xmuch 18

configs <- emo\_tb3 %>%  
 dplyr::select(Variety, Emotionality, Data)  
counts = emo\_tb3$Frequency  
cfa::cfa(configs, counts)

##   
## \*\*\* Analysis of configuration frequencies (CFA) \*\*\*  
##   
## label n expected Q chisq  
## 1 Australia emotional Xmuch 60 8.374729 0.024196034 318.2393642  
## 2 Ireland nonemotional Xmuch 18 60.680052 0.020506242 30.0195328  
## 3 Ireland emotional control 73 109.870528 0.018143789 12.3730708  
## 4 Ireland emotional Xmuch 15 7.800807 0.003373252 6.6439754  
## 5 Ireland nonemotional control 927 854.648613 0.056201739 6.1250005  
## 6 Australia emotional control 96 117.953935 0.010846559 4.0861314  
## 7 Australia nonemotional Xmuch 49 65.144412 0.007773488 4.0009882  
## 8 Australia nonemotional control 904 917.526924 0.011047138 0.1994248  
## p.chisq sig.chisq z p.z sig.z  
## 1 0.000000e+00 TRUE 17.7011258 0.0000000000 TRUE  
## 2 4.277161e-08 TRUE -5.6234197 0.9999999906 TRUE  
## 3 4.355706e-04 TRUE -3.6603505 0.9998740648 TRUE  
## 4 9.949168e-03 FALSE 2.4029511 0.0081316805 FALSE  
## 5 1.332833e-02 FALSE 3.1703159 0.0007613665 TRUE  
## 6 4.323650e-02 FALSE -2.1268447 0.9832835074 FALSE  
## 7 4.547359e-02 FALSE -2.0942882 0.9818828428 FALSE  
## 8 6.551855e-01 FALSE -0.6124747 0.7298881262 FALSE  
##   
##   
## Summary statistics:  
##   
## Total Chi squared = 381.6875   
## Total degrees of freedom = 4   
## p = 0   
## Sum of counts = 2142   
##   
## Levels:  
##   
## Variety Emotionality Data   
## 2 2 2

There are significantly more negative adjectives in the x-much construction in the Australian data after corrections compared to any other configuration.

### Visualization

* prepare data

emo\_vis <- emo\_tb3 %>%  
 dplyr::ungroup() %>%  
 dplyr::mutate\_if(is.character, factor) %>%  
 dplyr::mutate(Data = ifelse(Data == "Xmuch", "X-much", "control"))  
# inspect  
head(emo\_vis)

## # A tibble: 6 x 4  
## Emotionality Variety Data Frequency  
## <fct> <fct> <chr> <int>  
## 1 emotional Australia control 96  
## 2 emotional Australia X-much 60  
## 3 emotional Ireland control 73  
## 4 emotional Ireland X-much 15  
## 5 nonemotional Australia control 904  
## 6 nonemotional Australia X-much 49

emo1 <- emo\_vis %>%  
 dplyr::filter(Emotionality == "emotional") %>%  
 dplyr::pull()  
emo2 <- emo\_vis %>%  
 dplyr::filter(Emotionality == "nonemotional") %>%  
 dplyr::pull()  
# add dimnames  
column.names <- c("Australia", "Ireland")  
row.names <- c("control", "X-much")  
matrix.names <- c("emotional", "nonemotional")  
# generate matrix  
emo\_mx <- array(c(emo1, emo2), dim = c(2, 2, 2),  
 dimnames = list(row.names,   
 column.names,  
 matrix.names))  
# inspect  
emo\_mx

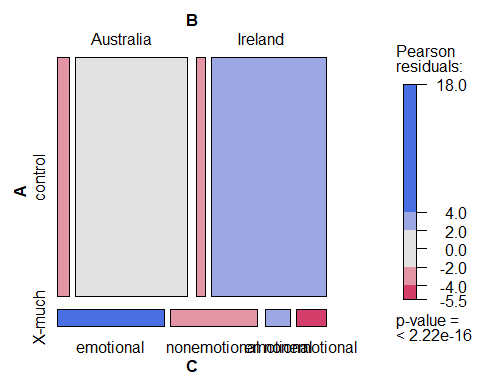
## , , emotional  
##   
## Australia Ireland  
## control 96 73  
## X-much 60 15  
##   
## , , nonemotional  
##   
## Australia Ireland  
## control 904 927  
## X-much 49 18

Generate mosaic plot

# open connection  
png(here::here("images", "emo\_mosaic.png"), width = 750, height = 300)  
# generate plot  
mosaic(emo\_mx,  
 axis.cex = 15,  
 shade = TRUE,  
 direction = c("h", "v", "v"),  
 just\_labels = c("center", "center", "center", "center"))  
# close window  
dev.off()

## png   
## 2

# show plot  
mosaic(emo\_mx,  
 shade = TRUE,  
 direction = c("h", "v", "v"),  
 just\_labels = c("center", "center", "center", "center"))



## Polarity

### X2 (wo variety)

RQ: Does the polarity of words in the x-much structure differ from the polarity in the control data?

Prepare data

pol\_tb1 <- xmuch %>%  
 dplyr::group\_by(Polarity, Status) %>%  
 dplyr::summarise(Frequency = n()) %>%  
 dplyr::ungroup() %>%  
 tidyr::spread(Status, Frequency) %>%  
 dplyr::mutate(phrase = replace\_na(phrase, 0)) %>%  
 tidyr::gather(Status, Frequency, phrase:word)

## `summarise()` has grouped output by 'Polarity'. You can override using the `.groups` argument.

# inspect  
pol\_tb1

## # A tibble: 6 x 3  
## Polarity Status Frequency  
## <chr> <chr> <dbl>  
## 1 negative phrase 15  
## 2 neutral phrase 26  
## 3 positive phrase 0  
## 4 negative word 102  
## 5 neutral word 1872  
## 6 positive word 127

Perform x2-test

pol\_x2 <- pol\_tb1 %>%  
 tidyr::spread(Status, Frequency) %>%  
 as.matrix()  
# inspect  
pol\_x2

## Polarity phrase word   
## [1,] "negative" "15" " 102"  
## [2,] "neutral" "26" "1872"  
## [3,] "positive" " 0" " 127"

rn <- pol\_x2[,1]  
pol\_x2 = pol\_x2[, 2:3]  
pol\_x2 <- apply(pol\_x2, 2, as.numeric)  
rownames(pol\_x2) <- rn  
pol\_x2

## phrase word  
## negative 15 102  
## neutral 26 1872  
## positive 0 127

# perform x2 test  
chisq.test(pol\_x2)

## Warning in chisq.test(pol\_x2): Chi-squared approximation may be incorrect

##   
## Pearson's Chi-squared test  
##   
## data: pol\_x2  
## X-squared = 79.6, df = 2, p-value < 2.2e-16

# effect size  
effectsize::effectsize(chisq.test(pol\_x2), type = "phi")

## Warning in chisq.test(pol\_x2): Chi-squared approximation may be incorrect

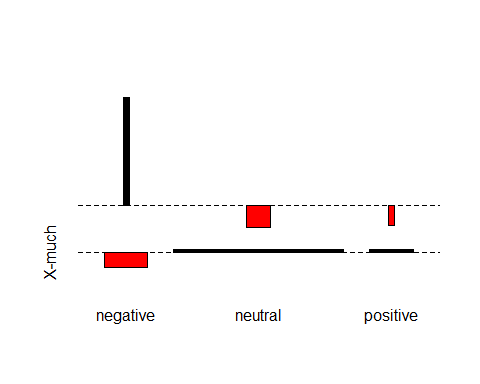
## Phi | 95% CI  
## -------------------  
## 0.19 | [0.16, 1.00]  
##   
## - One-sided CIs: upper bound fixed at (1).

The polarity of words in the X-much structure is significantly different from the polarity of words in the control data.

# adapt column names  
colnames(pol\_x2) <- c("control", "X-much")  
  
# open window  
png(here::here("images", "pol\_assoc.png"), width = 500, height = 350)  
# generate plot  
assocplot(pol\_x2)  
# close window  
dev.off()

## png   
## 2

assocplot(pol\_x2)



### X2 (wo variety)

RQ: Does the polarity of words in the x-much structure in Australia differ from the polarity of words in the x-much structure in Ireland?

Prepare data

pol\_tb2 <- xmuch %>%  
 dplyr::filter(Data == "Xmuch") %>%  
 dplyr::group\_by(Polarity, Variety) %>%  
 dplyr::summarise(Frequency = n())

## `summarise()` has grouped output by 'Polarity'. You can override using the `.groups` argument.

# inspect  
pol\_tb2

## # A tibble: 6 x 3  
## # Groups: Polarity [3]  
## Polarity Variety Frequency  
## <chr> <chr> <int>  
## 1 negative Australia 51  
## 2 negative Ireland 14  
## 3 neutral Australia 49  
## 4 neutral Ireland 18  
## 5 positive Australia 9  
## 6 positive Ireland 1

Perform x2-test

pol\_x22 <- pol\_tb2 %>%  
 tidyr::spread(Variety, Frequency) %>%  
 as.matrix()  
# inspect  
pol\_x22

## Polarity Australia Ireland  
## [1,] "negative" "51" "14"   
## [2,] "neutral" "49" "18"   
## [3,] "positive" " 9" " 1"

rn <- pol\_x22[,1]  
pol\_x22 = pol\_x22[, 2:3]  
pol\_x22 <- apply(pol\_x22, 2, as.numeric)  
rownames(pol\_x22) <- rn  
pol\_x22

## Australia Ireland  
## negative 51 14  
## neutral 49 18  
## positive 9 1

# perform x2 test  
chisq.test(pol\_x22)

## Warning in chisq.test(pol\_x22): Chi-squared approximation may be incorrect

##   
## Pearson's Chi-squared test  
##   
## data: pol\_x22  
## X-squared = 1.5819, df = 2, p-value = 0.4534

# effect size  
effectsize::effectsize(chisq.test(pol\_x22), type = "phi")

## Warning in chisq.test(pol\_x22): Chi-squared approximation may be incorrect

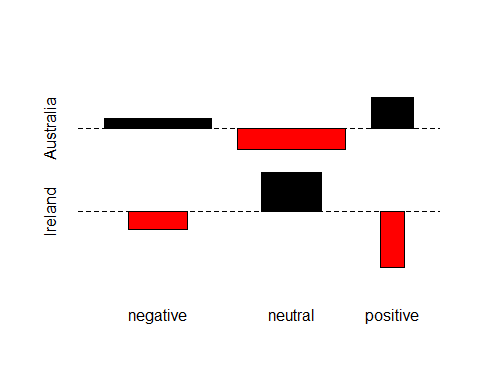
## Phi | 95% CI  
## -------------------  
## 0.11 | [0.00, 1.00]  
##   
## - One-sided CIs: upper bound fixed at (1).

Australia and Ireland do **not** differ significantly with respect to the polarity of the words in the X-much structure.

# open window  
png(here::here("images", "pol\_var\_assoc.png"), width = 500, height = 350)  
# generate plot  
assocplot(pol\_x22)  
# close window  
dev.off()

## png   
## 2

assocplot(pol\_x22)



### CFA

pol\_tb1 <- xmuch %>%  
 dplyr::group\_by(Polarity, Variety, Status) %>%  
 dplyr::summarise(Frequency = n())

## `summarise()` has grouped output by 'Polarity', 'Variety'. You can override using the `.groups` argument.

# inspect  
pol\_tb1

## # A tibble: 10 x 4  
## # Groups: Polarity, Variety [6]  
## Polarity Variety Status Frequency  
## <chr> <chr> <chr> <int>  
## 1 negative Australia phrase 11  
## 2 negative Australia word 66  
## 3 negative Ireland phrase 4  
## 4 negative Ireland word 36  
## 5 neutral Australia phrase 20  
## 6 neutral Australia word 933  
## 7 neutral Ireland phrase 6  
## 8 neutral Ireland word 939  
## 9 positive Australia word 79  
## 10 positive Ireland word 48

configs <- pol\_tb1 %>%  
 dplyr::select(Variety, Polarity, Status)  
counts = pol\_tb1$Frequency  
cfa::cfa(configs, counts)

##   
## \*\*\* Analysis of configuration frequencies (CFA) \*\*\*  
##   
## label n expected Q chisq p.chisq  
## 1 Australia negative phrase 11 1.159478 0.004596570 83.51682665 0.00000000  
## 2 Ireland negative phrase 4 1.080018 0.001363891 7.89458232 0.01930693  
## 3 Ireland neutral phrase 6 17.520297 0.005422644 7.57505631 0.02265152  
## 4 Ireland negative word 36 55.344351 0.009270505 6.76137535 0.03402405  
## 5 Australia positive word 79 64.494456 0.006982192 3.26246328 0.19568841  
## 6 Ireland positive word 48 60.074638 0.005799746 2.42692900 0.29716596  
## 7 Ireland neutral word 939 897.808368 0.033107144 1.88988050 0.38870280  
## 8 Australia neutral word 933 963.862033 0.026195602 0.98817576 0.61012716  
## 9 Australia negative word 66 59.416153 0.003161384 0.72954984 0.69435292  
## 10 Australia neutral phrase 20 18.809302 0.000560806 0.07537557 0.96301356  
## sig.chisq z p.z sig.z  
## 1 TRUE 9.1387541 0.000000000 TRUE  
## 2 FALSE 2.8097299 0.002479154 TRUE  
## 3 FALSE 2.7522820 0.002959077 TRUE  
## 4 FALSE 2.6002645 0.004657597 TRUE  
## 5 FALSE 1.8062290 0.035441287 FALSE  
## 6 FALSE 1.5578604 0.059633173 FALSE  
## 7 FALSE 1.3747292 0.084607700 FALSE  
## 8 FALSE 0.9940703 0.160094322 FALSE  
## 9 FALSE 0.8541369 0.196514571 FALSE  
## 10 FALSE 0.2745461 0.391832483 FALSE  
##   
##   
## Summary statistics:  
##   
## Total Chi squared = 115.1202   
## Total degrees of freedom = 7   
## p = 0   
## Sum of counts = 2142   
##   
## Levels:  
##   
## Variety Polarity Status   
## 2 3 2

### Visualization

* prepare data

pol\_vis <- pol\_tb1 %>%  
 dplyr::ungroup() %>%  
 dplyr::mutate\_if(is.character, factor)  
# inspect  
pol\_vis

## # A tibble: 10 x 4  
## Polarity Variety Status Frequency  
## <fct> <fct> <fct> <int>  
## 1 negative Australia phrase 11  
## 2 negative Australia word 66  
## 3 negative Ireland phrase 4  
## 4 negative Ireland word 36  
## 5 neutral Australia phrase 20  
## 6 neutral Australia word 933  
## 7 neutral Ireland phrase 6  
## 8 neutral Ireland word 939  
## 9 positive Australia word 79  
## 10 positive Ireland word 48

* convert to matrix

pol1 <- pol\_vis %>%  
 dplyr::filter(Polarity == "positive") %>%  
 dplyr::pull(Frequency)  
pol2 <- pol\_vis %>%  
 dplyr::filter(Polarity == "neutral") %>%  
 dplyr::pull(Frequency)  
pol3 <- pol\_vis %>%  
 dplyr::filter(Polarity == "negative") %>%  
 dplyr::pull(Frequency)  
# add dimnames  
column.names <- c("Australia", "Ireland")  
row.names <- c("control", "X-much")  
matrix.names <- c("positive", "neutral", "negative")  
# generate matrix  
pol\_mx <- array(c(pol1, pol2, pol3), dim = c(2, 2, 3),  
 dimnames = list(row.names,   
 column.names,  
 matrix.names))  
# inspect  
pol\_mx

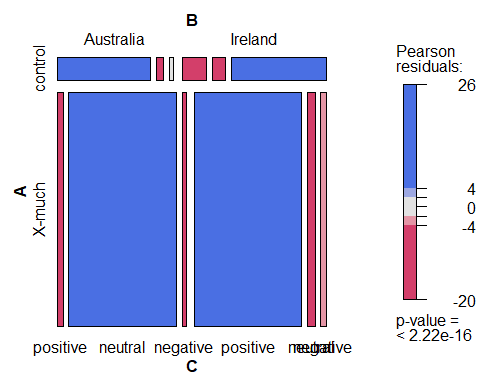
## , , positive  
##   
## Australia Ireland  
## control 79 20  
## X-much 48 933  
##   
## , , neutral  
##   
## Australia Ireland  
## control 6 11  
## X-much 939 66  
##   
## , , negative  
##   
## Australia Ireland  
## control 4 79  
## X-much 36 48

* mosaic plot

# open window  
png(here::here("images", "pol\_mosaic.png"), width = 750, height = 300)  
# generate plot  
mosaic(pol\_mx,  
 shade = TRUE,  
 direction = c("h", "v", "v"),  
 just\_labels = c("center", "center", "center", "center"))  
# close window  
dev.off()

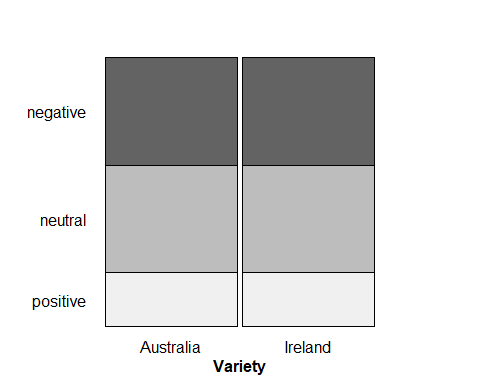
## png   
## 2

# show plot  
mosaic(pol\_mx,  
 shade = TRUE,  
 direction = c("h", "v", "v"),  
 just\_labels = c("center", "center", "center", "center"))



* alternative mosaic plot

# define cols  
mycols <- rev(RColorBrewer::brewer.pal(3, "Greys"))  
vcd::mosaic(Polarity ~ Variety, data = pol\_vis,   
 direction = c("v", "h"),  
 rot\_labels = c(0, 0, 0, 0),  
 highlighting\_fill = mycols,  
 labeling\_args = list(tl\_labels = c(F, T),  
 set\_varnames = c(Polarity = "Polarity\n\n\n\n\n\n\n"),  
 offset\_varnames = c(0, 0, 0, 1),  
 just\_labels = c("center", "center", "center", "right")))



# save plot  
#ggsave(here::here("images", "pol\_var.png"), units = "cm", width = 8, height = 6)

## POS

tabulation

pos\_tab <- xmuch %>%  
 dplyr::group\_by(Variety, Data, POS) %>%  
 dplyr::summarise(Freq= n()) %>%  
 dplyr::group\_by(Variety, Data) %>%  
 dplyr::mutate(Total = sum(Freq)) %>%  
 dplyr::rowwise() %>%  
 dplyr::mutate(Percent = round(Freq/Total\*100, 1),  
 Frequency = paste0(Freq, " (", Percent, ")")) %>%  
 dplyr::ungroup() %>%  
 dplyr::select(-Freq, -Total, -Percent) %>%  
 tidyr::spread(POS, Frequency)

## `summarise()` has grouped output by 'Variety', 'Data'. You can override using the `.groups` argument.

# inspect  
pos\_tab

## # A tibble: 4 x 6  
## Variety Data Adjective Noun other Verb   
## <chr> <chr> <chr> <chr> <chr> <chr>   
## 1 Australia control 75 (7.5) 322 (32.2) 526 (52.6) 77 (7.7)   
## 2 Australia Xmuch 45 (41.3) 46 (42.2) 2 (1.8) 16 (14.7)  
## 3 Ireland control 79 (7.9) 316 (31.6) 543 (54.3) 62 (6.2)   
## 4 Ireland Xmuch 8 (24.2) 23 (69.7) 1 (3) 1 (3)

RQ: Do the word classes (parts-of-speech) of words in the x-much structure differ from the word classes in the control data?

### X2 (wo variety)

Prepare data

pos\_tb1 <- xmuch %>%  
 dplyr::group\_by(POS, Data) %>%  
 dplyr::summarise(Frequency = n())

## `summarise()` has grouped output by 'POS'. You can override using the `.groups` argument.

# inspect  
pos\_tb1

## # A tibble: 8 x 3  
## # Groups: POS [4]  
## POS Data Frequency  
## <chr> <chr> <int>  
## 1 Adjective control 154  
## 2 Adjective Xmuch 53  
## 3 Noun control 638  
## 4 Noun Xmuch 69  
## 5 other control 1069  
## 6 other Xmuch 3  
## 7 Verb control 139  
## 8 Verb Xmuch 17

Perform x2-test

pos\_x2 <- pos\_tb1 %>%  
 tidyr::spread(Data, Frequency) %>%  
 as.matrix()  
# inspect  
pos\_x2

## POS control Xmuch  
## [1,] "Adjective" " 154" "53"   
## [2,] "Noun" " 638" "69"   
## [3,] "other" "1069" " 3"   
## [4,] "Verb" " 139" "17"

rn <- pos\_x2[,1]  
pos\_x2 = pos\_x2[, 2:3]  
pos\_x2 <- apply(pos\_x2, 2, as.numeric)  
rownames(pos\_x2) <- rn  
pos\_x2

## control Xmuch  
## Adjective 154 53  
## Noun 638 69  
## other 1069 3  
## Verb 139 17

# perform x2 test  
chisq.test(pos\_x2)

##   
## Pearson's Chi-squared test  
##   
## data: pos\_x2  
## X-squared = 206.01, df = 3, p-value < 2.2e-16

# effect size  
effectsize::effectsize(chisq.test(pos\_x2), type = "phi")

## Phi | 95% CI  
## -------------------  
## 0.31 | [0.27, 1.00]  
##   
## - One-sided CIs: upper bound fixed at (1).

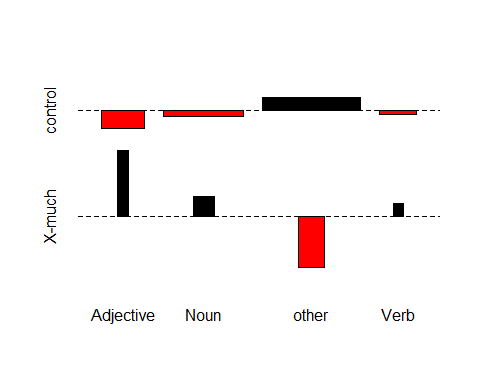
The word classes (parts-of-speech) of words in the x-much structure differ significantly from the word classes in the control data!

generate association plot

# adapt column names  
colnames(pos\_x2) <- c("control", "X-much")  
  
# open window  
png(here::here("images", "pos\_assoc.png"), width = 500, height = 350)  
# generate plot  
assocplot(pos\_x2)  
# close window  
dev.off()

## png   
## 2

assocplot(pos\_x2)



### X2 (variety)

RQ: Do Australia and Ireland differ with respect to the word classes (parts-of-speech) of words in the x-much structure?

Prepare data

pos\_tb2 <- xmuch %>%  
 dplyr::filter(Data == "Xmuch") %>%  
 dplyr::group\_by(POS, Variety) %>%  
 dplyr::summarise(Frequency = n())

## `summarise()` has grouped output by 'POS'. You can override using the `.groups` argument.

# inspect  
pos\_tb2

## # A tibble: 8 x 3  
## # Groups: POS [4]  
## POS Variety Frequency  
## <chr> <chr> <int>  
## 1 Adjective Australia 45  
## 2 Adjective Ireland 8  
## 3 Noun Australia 46  
## 4 Noun Ireland 23  
## 5 other Australia 2  
## 6 other Ireland 1  
## 7 Verb Australia 16  
## 8 Verb Ireland 1

Perform x2-test

pos\_x22 <- pos\_tb2 %>%  
 tidyr::spread(Variety, Frequency) %>%  
 as.matrix()  
# inspect  
pos\_x22

## POS Australia Ireland  
## [1,] "Adjective" "45" " 8"   
## [2,] "Noun" "46" "23"   
## [3,] "other" " 2" " 1"   
## [4,] "Verb" "16" " 1"

rn <- pos\_x22[,1]  
pos\_x22 = pos\_x22[, 2:3]  
pos\_x22 <- apply(pos\_x22, 2, as.numeric)  
rownames(pos\_x22) <- rn  
pos\_x22

## Australia Ireland  
## Adjective 45 8  
## Noun 46 23  
## other 2 1  
## Verb 16 1

# perform x2 test  
chisq.test(pos\_x22)

## Warning in chisq.test(pos\_x22): Chi-squared approximation may be incorrect

##   
## Pearson's Chi-squared test  
##   
## data: pos\_x22  
## X-squared = 8.9544, df = 3, p-value = 0.0299

# effect size  
effectsize::effectsize(chisq.test(pos\_x22), type = "phi")

## Warning in chisq.test(pos\_x22): Chi-squared approximation may be incorrect

## Phi | 95% CI  
## -------------------  
## 0.25 | [0.06, 1.00]  
##   
## - One-sided CIs: upper bound fixed at (1).

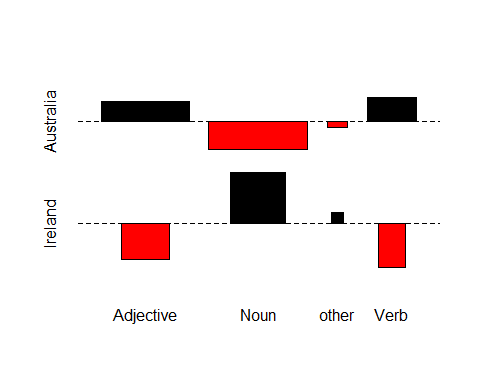
Australia and Ireland differ significantly with respect to the word classes (parts-of-speech) of words in the x-much structure!

generate association plot

# open window  
png(here::here("images", "pos\_var\_assoc.png"), width = 500, height = 350)  
# generate plot  
assocplot(pos\_x22)  
# close window  
dev.off()

## png   
## 2

assocplot(pos\_x22)



### CFA

process data

pos\_tb2 <- xmuch %>%  
 dplyr::group\_by(POS, Variety, Data) %>%  
 dplyr::summarise(Frequency = n())

## `summarise()` has grouped output by 'POS', 'Variety'. You can override using the `.groups` argument.

# inspect  
pos\_tb2

## # A tibble: 16 x 4  
## # Groups: POS, Variety [8]  
## POS Variety Data Frequency  
## <chr> <chr> <chr> <int>  
## 1 Adjective Australia control 75  
## 2 Adjective Australia Xmuch 45  
## 3 Adjective Ireland control 79  
## 4 Adjective Ireland Xmuch 8  
## 5 Noun Australia control 322  
## 6 Noun Australia Xmuch 46  
## 7 Noun Ireland control 316  
## 8 Noun Ireland Xmuch 23  
## 9 other Australia control 526  
## 10 other Australia Xmuch 2  
## 11 other Ireland control 543  
## 12 other Ireland Xmuch 1  
## 13 Verb Australia control 77  
## 14 Verb Australia Xmuch 16  
## 15 Verb Ireland control 62  
## 16 Verb Ireland Xmuch 1

perform CFA

configs <- pos\_tb2 %>%  
 dplyr::select(Variety, POS, Data)  
counts = pos\_tb2$Frequency  
cfa::cfa(configs, counts)

##   
## \*\*\* Analysis of configuration frequencies (CFA) \*\*\*  
##   
## label n expected Q chisq  
## 1 Australia Adjective Xmuch 45 7.104791 0.0177503837 2.021237e+02  
## 2 Australia other Xmuch 2 36.793893 0.0165275471 3.290261e+01  
## 3 Ireland other Xmuch 1 34.272400 0.0157859109 3.230158e+01  
## 4 Australia Verb Xmuch 16 5.354335 0.0049824194 2.116606e+01  
## 5 Australia Noun Xmuch 46 24.266122 0.0102627996 1.946588e+01  
## 6 Ireland other control 543 482.709860 0.0363348995 7.530198e+00  
## 7 Australia Adjective control 75 100.067478 0.0122763498 6.279547e+00  
## 8 Ireland Verb Xmuch 1 4.987401 0.0018658765 3.187907e+00  
## 9 Ireland Adjective control 79 93.209833 0.0069357190 2.166288e+00  
## 10 Australia Noun control 322 341.776362 0.0109855029 1.144329e+00  
## 11 Ireland Verb control 62 70.245092 0.0039797620 9.677763e-01  
## 12 Ireland Adjective Xmuch 8 6.617898 0.0006472387 2.886423e-01  
## 13 Australia other control 526 518.223847 0.0047889315 1.166842e-01  
## 14 Australia Verb control 77 75.413172 0.0007678498 3.338971e-02  
## 15 Ireland Noun control 316 318.354357 0.0012910167 1.741141e-02  
## 16 Ireland Noun Xmuch 23 22.603159 0.0001872423 6.967278e-03  
## p.chisq sig.chisq z p.z sig.z  
## 1 0.000000e+00 TRUE 14.21702267 0.000000e+00 TRUE  
## 2 3.376675e-07 TRUE 5.73607940 4.844663e-09 TRUE  
## 3 4.520883e-07 TRUE 5.68344772 6.600304e-09 TRUE  
## 4 9.723848e-05 TRUE 4.60065863 2.105786e-06 TRUE  
## 5 2.189855e-04 TRUE 4.41201566 5.120635e-06 TRUE  
## 6 5.678758e-02 FALSE 2.74412071 3.033661e-03 TRUE  
## 7 9.877406e-02 FALSE 2.50590247 6.106965e-03 FALSE  
## 8 3.635511e-01 FALSE 1.78547100 3.709248e-02 FALSE  
## 9 5.386191e-01 FALSE 1.47183155 7.053319e-02 FALSE  
## 10 7.663852e-01 FALSE 1.06973299 1.423698e-01 FALSE  
## 11 8.090485e-01 FALSE 0.98375621 1.626177e-01 FALSE  
## 12 9.621502e-01 FALSE 0.53725444 2.955459e-01 FALSE  
## 13 9.897627e-01 FALSE 0.34159074 3.663295e-01 FALSE  
## 14 9.983935e-01 FALSE 0.18272851 4.275055e-01 FALSE  
## 15 9.993921e-01 FALSE 0.13195229 4.475110e-01 FALSE  
## 16 9.998457e-01 FALSE 0.08347022 4.667388e-01 FALSE  
##   
##   
## Summary statistics:  
##   
## Total Chi squared = 329.699   
## Total degrees of freedom = 10   
## p = 0   
## Sum of counts = 2142   
##   
## Levels:  
##   
## Variety POS Data   
## 2 4 2

### Visualization

* prepare data

pos\_vis <- pos\_tb2 %>%  
 dplyr::ungroup() %>%  
 dplyr::mutate\_if(is.character, factor)  
# inspect  
pos\_vis

## # A tibble: 16 x 4  
## POS Variety Data Frequency  
## <fct> <fct> <fct> <int>  
## 1 Adjective Australia control 75  
## 2 Adjective Australia Xmuch 45  
## 3 Adjective Ireland control 79  
## 4 Adjective Ireland Xmuch 8  
## 5 Noun Australia control 322  
## 6 Noun Australia Xmuch 46  
## 7 Noun Ireland control 316  
## 8 Noun Ireland Xmuch 23  
## 9 other Australia control 526  
## 10 other Australia Xmuch 2  
## 11 other Ireland control 543  
## 12 other Ireland Xmuch 1  
## 13 Verb Australia control 77  
## 14 Verb Australia Xmuch 16  
## 15 Verb Ireland control 62  
## 16 Verb Ireland Xmuch 1

* prepare data

pos1 <- pos\_vis %>%  
 dplyr::filter(POS == "Adjective") %>%  
 dplyr::pull(Frequency)  
pos2 <- pos\_vis %>%  
 dplyr::filter(POS == "Noun") %>%  
 dplyr::pull(Frequency)  
pos3 <- pos\_vis %>%  
 dplyr::filter(POS == "other") %>%  
 dplyr::pull(Frequency)  
pos4 <- pos\_vis %>%  
 dplyr::filter(POS == "Verb") %>%  
 dplyr::pull(Frequency)  
# add dimnames  
column.names <- c("Australia", "Ireland")  
row.names <- c("control", "X-much")  
matrix.names <- c("Adj.", "Noun", "other", "Verb")  
# generate matrix  
pos\_mx <- array(c(pol1, pol2, pol3), dim = c(2, 2, 4),  
 dimnames = list(row.names,   
 column.names,  
 matrix.names))  
# inspect  
pos\_mx

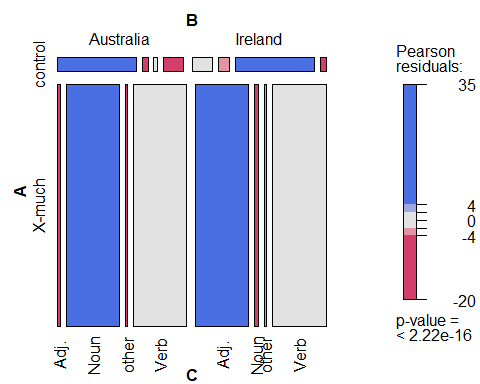
## , , Adj.  
##   
## Australia Ireland  
## control 79 20  
## X-much 48 933  
##   
## , , Noun  
##   
## Australia Ireland  
## control 6 11  
## X-much 939 66  
##   
## , , other  
##   
## Australia Ireland  
## control 4 79  
## X-much 36 48  
##   
## , , Verb  
##   
## Australia Ireland  
## control 20 6  
## X-much 933 939

Generate mosaic plot

# open window  
png(here::here("images", "pos\_mosaic.png"), width = 750, height = 300)  
# generate plot  
mosaic(pos\_mx,  
 shade = TRUE,  
 direction = c("h", "v", "v"),  
 just\_labels = c("center", "center", "center", "center"),  
 rot\_labels=c(0,0,90,90),  
 offset\_labels = c(0, 0, 0.5,0))  
# close window  
dev.off()

## png   
## 2

# show plot  
mosaic(pos\_mx,  
 shade = TRUE,  
 direction = c("h", "v", "v"),  
 just\_labels = c("center", "center", "center", "center"),  
 rot\_labels=c(0,0,90,90),  
 offset\_labels = c(0, 0, 0.5,0))



## Words

RQ: Are there Words that are significantly over-represented in the X-much structure?

**Constructionalization more advanced in AusE**

### CFA (wo vareity)

RQ: Regardless of variety, are there words that are significantly attracted by the X-much structure?

wordcat\_tb <- xmuch %>%  
 dplyr::group\_by(Data, X) %>%  
 dplyr::mutate(Frequency = n()) %>%  
 dplyr::mutate(X = ifelse(Frequency < 2, "other", X)) %>%  
 dplyr::ungroup() %>%  
 dplyr::group\_by(X, Data) %>%  
 dplyr::summarise(Frequency = n())

## `summarise()` has grouped output by 'X'. You can override using the `.groups` argument.

configs <- wordcat\_tb %>%  
 dplyr::select(Data, X)  
counts = wordcat\_tb$Frequency  
cfa::cfa(configs, counts)

##   
## \*\*\* Analysis of configuration frequencies (CFA) \*\*\*  
##   
## label n expected Q chisq p.chisq  
## 1 Xmuch jealous 7 0.4647031 3.055969e-03 91.908365104 1  
## 2 Xmuch paranoid 4 0.2655446 1.746105e-03 52.519065774 1  
## 3 Xmuch agenda 4 0.2655446 1.746105e-03 52.519065774 1  
## 4 Xmuch threatened 3 0.1991585 1.309538e-03 39.389299330 1  
## 5 Xmuch other 104 58.5525947 2.184502e-02 35.275407739 1  
## 6 Xmuch hypocrite 2 0.1327723 8.729984e-04 26.259532887 1  
## 7 Xmuch hipster 2 0.1327723 8.729984e-04 26.259532887 1  
## 8 Xmuch generalise 2 0.1327723 8.729984e-04 26.259532887 1  
## 9 Xmuch excuses 2 0.1327723 8.729984e-04 26.259532887 1  
## 10 Xmuch excited 2 0.1327723 8.729984e-04 26.259532887 1  
## 11 Xmuch entitled 2 0.1327723 8.729984e-04 26.259532887 1  
## 12 Xmuch creepy 2 0.1327723 8.729984e-04 26.259532887 1  
## 13 Xmuch coincidence 2 0.1327723 8.729984e-04 26.259532887 1  
## 14 Xmuch bitter 2 0.1327723 8.729984e-04 26.259532887 1  
## 15 Xmuch biased 2 0.1327723 8.729984e-04 26.259532887 1  
## 16 control other 778 823.4474053 3.454625e-02 2.508316424 1  
## 17 control the 123 114.8345021 4.034007e-03 0.580621283 1  
## 18 control and 53 49.4815334 1.683865e-03 0.250186406 1  
## 19 control a 53 49.4815334 1.683865e-03 0.250186406 1  
## 20 control to 51 47.6143058 1.618876e-03 0.240745410 1  
## 21 control in 51 47.6143058 1.618876e-03 0.240745410 1  
## 22 control of 41 38.2781674 1.295665e-03 0.193540428 1  
## 23 control is 33 30.8092567 1.039158e-03 0.155776442 1  
## 24 control i 23 21.4731183 7.210684e-04 0.108571459 1  
## 25 control for 23 21.4731183 7.210684e-04 0.108571459 1  
## 26 control was 21 19.6058906 6.577868e-04 0.099130463 1  
## 27 control on 17 15.8714353 5.315574e-04 0.080248470 1  
## 28 control are 17 15.8714353 5.315574e-04 0.080248470 1  
## 29 control you 16 14.9378214 5.000694e-04 0.075527972 1  
## 30 control that 15 14.0042076 4.686091e-04 0.070807473 1  
## 31 control as 15 14.0042076 4.686091e-04 0.070807473 1  
## 32 control not 14 13.0705937 4.371764e-04 0.066086975 1  
## 33 control it 14 13.0705937 4.371764e-04 0.066086975 1  
## 34 control with 13 12.1369799 4.057714e-04 0.061366477 1  
## 35 control from 12 11.2033661 3.743938e-04 0.056645979 1  
## 36 control he 11 10.2697522 3.430438e-04 0.051925481 1  
## 37 control be 11 10.2697522 3.430438e-04 0.051925481 1  
## 38 control this 10 9.3361384 3.117213e-04 0.047204982 1  
## 39 control they 10 9.3361384 3.117213e-04 0.047204982 1  
## 40 control there 10 9.3361384 3.117213e-04 0.047204982 1  
## 41 control at 9 8.4025245 2.804262e-04 0.042484484 1  
## 42 control an 9 8.4025245 2.804262e-04 0.042484484 1  
## 43 control which 8 7.4689107 2.491586e-04 0.037763986 1  
## 44 control so 8 7.4689107 2.491586e-04 0.037763986 1  
## 45 control or 8 7.4689107 2.491586e-04 0.037763986 1  
## 46 control no 8 7.4689107 2.491586e-04 0.037763986 1  
## 47 control have 8 7.4689107 2.491586e-04 0.037763986 1  
## 48 control do 8 7.4689107 2.491586e-04 0.037763986 1  
## 49 control would 7 6.5352969 2.179183e-04 0.033043488 1  
## 50 control up 7 6.5352969 2.179183e-04 0.033043488 1  
## 51 control their 7 6.5352969 2.179183e-04 0.033043488 1  
## 52 control some 7 6.5352969 2.179183e-04 0.033043488 1  
## 53 control out 7 6.5352969 2.179183e-04 0.033043488 1  
## 54 control my 7 6.5352969 2.179183e-04 0.033043488 1  
## 55 control if 7 6.5352969 2.179183e-04 0.033043488 1  
## 56 control documentary 7 6.5352969 2.179183e-04 0.033043488 1  
## 57 control can 7 6.5352969 2.179183e-04 0.033043488 1  
## 58 control but 7 6.5352969 2.179183e-04 0.033043488 1  
## 59 control about 7 6.5352969 2.179183e-04 0.033043488 1  
## 60 control who 6 5.6016830 1.867054e-04 0.028322989 1  
## 61 control what 6 5.6016830 1.867054e-04 0.028322989 1  
## 62 control time 6 5.6016830 1.867054e-04 0.028322989 1  
## 63 control may 6 5.6016830 1.867054e-04 0.028322989 1  
## 64 control like 6 5.6016830 1.867054e-04 0.028322989 1  
## 65 control has 6 5.6016830 1.867054e-04 0.028322989 1  
## 66 control by 6 5.6016830 1.867054e-04 0.028322989 1  
## 67 control your 5 4.6680692 1.555198e-04 0.023602491 1  
## 68 control we 5 4.6680692 1.555198e-04 0.023602491 1  
## 69 control way 5 4.6680692 1.555198e-04 0.023602491 1  
## 70 control still 5 4.6680692 1.555198e-04 0.023602491 1  
## 71 control our 5 4.6680692 1.555198e-04 0.023602491 1  
## 72 control much 5 4.6680692 1.555198e-04 0.023602491 1  
## 73 control his 5 4.6680692 1.555198e-04 0.023602491 1  
## 74 control get 5 4.6680692 1.555198e-04 0.023602491 1  
## 75 control any 5 4.6680692 1.555198e-04 0.023602491 1  
## 76 control want 4 3.7344554 1.243614e-04 0.018881993 1  
## 77 control us 4 3.7344554 1.243614e-04 0.018881993 1  
## 78 control think 4 3.7344554 1.243614e-04 0.018881993 1  
## 79 control new 4 3.7344554 1.243614e-04 0.018881993 1  
## 80 control had 4 3.7344554 1.243614e-04 0.018881993 1  
## 81 control community 4 3.7344554 1.243614e-04 0.018881993 1  
## 82 control being 4 3.7344554 1.243614e-04 0.018881993 1  
## 83 control before 4 3.7344554 1.243614e-04 0.018881993 1  
## 84 control years 3 2.8008415 9.323030e-05 0.014161495 1  
## 85 control within 3 2.8008415 9.323030e-05 0.014161495 1  
## 86 control will 3 2.8008415 9.323030e-05 0.014161495 1  
## 87 control were 3 2.8008415 9.323030e-05 0.014161495 1  
## 88 control very 3 2.8008415 9.323030e-05 0.014161495 1  
## 89 control two 3 2.8008415 9.323030e-05 0.014161495 1  
## 90 control though 3 2.8008415 9.323030e-05 0.014161495 1  
## 91 control these 3 2.8008415 9.323030e-05 0.014161495 1  
## 92 control state 3 2.8008415 9.323030e-05 0.014161495 1  
## 93 control services 3 2.8008415 9.323030e-05 0.014161495 1  
## 94 control said 3 2.8008415 9.323030e-05 0.014161495 1  
## 95 control rt 3 2.8008415 9.323030e-05 0.014161495 1  
## 96 control people 3 2.8008415 9.323030e-05 0.014161495 1  
## 97 control part 3 2.8008415 9.323030e-05 0.014161495 1  
## 98 control one 3 2.8008415 9.323030e-05 0.014161495 1  
## 99 control more 3 2.8008415 9.323030e-05 0.014161495 1  
## 100 control months 3 2.8008415 9.323030e-05 0.014161495 1  
## 101 control lead 3 2.8008415 9.323030e-05 0.014161495 1  
## 102 control just 3 2.8008415 9.323030e-05 0.014161495 1  
## 103 control into 3 2.8008415 9.323030e-05 0.014161495 1  
## 104 control how 3 2.8008415 9.323030e-05 0.014161495 1  
## 105 control hours 3 2.8008415 9.323030e-05 0.014161495 1  
## 106 control government 3 2.8008415 9.323030e-05 0.014161495 1  
## 107 control god 3 2.8008415 9.323030e-05 0.014161495 1  
## 108 control found 3 2.8008415 9.323030e-05 0.014161495 1  
## 109 control early 3 2.8008415 9.323030e-05 0.014161495 1  
## 110 control different 3 2.8008415 9.323030e-05 0.014161495 1  
## 111 control day 3 2.8008415 9.323030e-05 0.014161495 1  
## 112 control business 3 2.8008415 9.323030e-05 0.014161495 1  
## 113 control bored 3 2.8008415 9.323030e-05 0.014161495 1  
## 114 control also 3 2.8008415 9.323030e-05 0.014161495 1  
## 115 control all 3 2.8008415 9.323030e-05 0.014161495 1  
## 116 control after 3 2.8008415 9.323030e-05 0.014161495 1  
## 117 control 2011 3 2.8008415 9.323030e-05 0.014161495 1  
## 118 control young 2 1.8672277 6.212638e-05 0.009440996 1  
## 119 control world 2 1.8672277 6.212638e-05 0.009440996 1  
## 120 control work 2 1.8672277 6.212638e-05 0.009440996 1  
## 121 control while 2 1.8672277 6.212638e-05 0.009440996 1  
## 122 control where 2 1.8672277 6.212638e-05 0.009440996 1  
## 123 control well 2 1.8672277 6.212638e-05 0.009440996 1  
## 124 control ways 2 1.8672277 6.212638e-05 0.009440996 1  
## 125 control water 2 1.8672277 6.212638e-05 0.009440996 1  
## 126 control vulnerable 2 1.8672277 6.212638e-05 0.009440996 1  
## 127 control unless 2 1.8672277 6.212638e-05 0.009440996 1  
## 128 control unique 2 1.8672277 6.212638e-05 0.009440996 1  
## 129 control types 2 1.8672277 6.212638e-05 0.009440996 1  
## 130 control turn 2 1.8672277 6.212638e-05 0.009440996 1  
## 131 control town 2 1.8672277 6.212638e-05 0.009440996 1  
## 132 control total 2 1.8672277 6.212638e-05 0.009440996 1  
## 133 control too 2 1.8672277 6.212638e-05 0.009440996 1  
## 134 control those 2 1.8672277 6.212638e-05 0.009440996 1  
## 135 control thanks 2 1.8672277 6.212638e-05 0.009440996 1  
## 136 control system 2 1.8672277 6.212638e-05 0.009440996 1  
## 137 control such 2 1.8672277 6.212638e-05 0.009440996 1  
## 138 control stay 2 1.8672277 6.212638e-05 0.009440996 1  
## 139 control school 2 1.8672277 6.212638e-05 0.009440996 1  
## 140 control return 2 1.8672277 6.212638e-05 0.009440996 1  
## 141 control really 2 1.8672277 6.212638e-05 0.009440996 1  
## 142 control race 2 1.8672277 6.212638e-05 0.009440996 1  
## 143 control put 2 1.8672277 6.212638e-05 0.009440996 1  
## 144 control position 2 1.8672277 6.212638e-05 0.009440996 1  
## 145 control politics 2 1.8672277 6.212638e-05 0.009440996 1  
## 146 control played 2 1.8672277 6.212638e-05 0.009440996 1  
## 147 control paragraph 2 1.8672277 6.212638e-05 0.009440996 1  
## 148 control over 2 1.8672277 6.212638e-05 0.009440996 1  
## 149 control outing 2 1.8672277 6.212638e-05 0.009440996 1  
## 150 control number 2 1.8672277 6.212638e-05 0.009440996 1  
## 151 control music 2 1.8672277 6.212638e-05 0.009440996 1  
## 152 control most 2 1.8672277 6.212638e-05 0.009440996 1  
## 153 control model 2 1.8672277 6.212638e-05 0.009440996 1  
## 154 control minutes 2 1.8672277 6.212638e-05 0.009440996 1  
## 155 control minimum 2 1.8672277 6.212638e-05 0.009440996 1  
## 156 control memorable 2 1.8672277 6.212638e-05 0.009440996 1  
## 157 control me 2 1.8672277 6.212638e-05 0.009440996 1  
## 158 control many 2 1.8672277 6.212638e-05 0.009440996 1  
## 159 control making 2 1.8672277 6.212638e-05 0.009440996 1  
## 160 control make 2 1.8672277 6.212638e-05 0.009440996 1  
## 161 control made 2 1.8672277 6.212638e-05 0.009440996 1  
## 162 control location 2 1.8672277 6.212638e-05 0.009440996 1  
## 163 control little 2 1.8672277 6.212638e-05 0.009440996 1  
## 164 control life 2 1.8672277 6.212638e-05 0.009440996 1  
## 165 control known 2 1.8672277 6.212638e-05 0.009440996 1  
## 166 control know 2 1.8672277 6.212638e-05 0.009440996 1  
## 167 control james 2 1.8672277 6.212638e-05 0.009440996 1  
## 168 control ireland 2 1.8672277 6.212638e-05 0.009440996 1  
## 169 control include 2 1.8672277 6.212638e-05 0.009440996 1  
## 170 control impressive 2 1.8672277 6.212638e-05 0.009440996 1  
## 171 control important 2 1.8672277 6.212638e-05 0.009440996 1  
## 172 control http 2 1.8672277 6.212638e-05 0.009440996 1  
## 173 control home 2 1.8672277 6.212638e-05 0.009440996 1  
## 174 control her 2 1.8672277 6.212638e-05 0.009440996 1  
## 175 control hair 2 1.8672277 6.212638e-05 0.009440996 1  
## 176 control good 2 1.8672277 6.212638e-05 0.009440996 1  
## 177 control go 2 1.8672277 6.212638e-05 0.009440996 1  
## 178 control further 2 1.8672277 6.212638e-05 0.009440996 1  
## 179 control four 2 1.8672277 6.212638e-05 0.009440996 1  
## 180 control following 2 1.8672277 6.212638e-05 0.009440996 1  
## 181 control first 2 1.8672277 6.212638e-05 0.009440996 1  
## 182 control find 2 1.8672277 6.212638e-05 0.009440996 1  
## 183 control fact 2 1.8672277 6.212638e-05 0.009440996 1  
## 184 control every 2 1.8672277 6.212638e-05 0.009440996 1  
## 185 control else 2 1.8672277 6.212638e-05 0.009440996 1  
## 186 control ear 2 1.8672277 6.212638e-05 0.009440996 1  
## 187 control down 2 1.8672277 6.212638e-05 0.009440996 1  
## 188 control director 2 1.8672277 6.212638e-05 0.009440996 1  
## 189 control didn 2 1.8672277 6.212638e-05 0.009440996 1  
## 190 control did 2 1.8672277 6.212638e-05 0.009440996 1  
## 191 control current 2 1.8672277 6.212638e-05 0.009440996 1  
## 192 control country 2 1.8672277 6.212638e-05 0.009440996 1  
## 193 control cork 2 1.8672277 6.212638e-05 0.009440996 1  
## 194 control computer 2 1.8672277 6.212638e-05 0.009440996 1  
## 195 control common 2 1.8672277 6.212638e-05 0.009440996 1  
## 196 control comment 2 1.8672277 6.212638e-05 0.009440996 1  
## 197 control come 2 1.8672277 6.212638e-05 0.009440996 1  
## 198 control church 2 1.8672277 6.212638e-05 0.009440996 1  
## 199 control brand 2 1.8672277 6.212638e-05 0.009440996 1  
## 200 control bill 2 1.8672277 6.212638e-05 0.009440996 1  
## 201 control begin 2 1.8672277 6.212638e-05 0.009440996 1  
## 202 control been 2 1.8672277 6.212638e-05 0.009440996 1  
## 203 control because 2 1.8672277 6.212638e-05 0.009440996 1  
## 204 control beautiful 2 1.8672277 6.212638e-05 0.009440996 1  
## 205 control bear 2 1.8672277 6.212638e-05 0.009440996 1  
## 206 control bank 2 1.8672277 6.212638e-05 0.009440996 1  
## 207 control ball 2 1.8672277 6.212638e-05 0.009440996 1  
## 208 control back 2 1.8672277 6.212638e-05 0.009440996 1  
## 209 control australia 2 1.8672277 6.212638e-05 0.009440996 1  
## 210 control ask 2 1.8672277 6.212638e-05 0.009440996 1  
## 211 control another 2 1.8672277 6.212638e-05 0.009440996 1  
## 212 control aid 2 1.8672277 6.212638e-05 0.009440996 1  
## 213 control adam 2 1.8672277 6.212638e-05 0.009440996 1  
## 214 control act 2 1.8672277 6.212638e-05 0.009440996 1  
## 215 control academic 2 1.8672277 6.212638e-05 0.009440996 1  
## 216 control 4 2 1.8672277 6.212638e-05 0.009440996 1  
## sig.chisq z p.z sig.z  
## 1 FALSE 9.58688506 0.000000e+00 TRUE  
## 2 FALSE 7.24700392 2.130518e-13 TRUE  
## 3 FALSE 7.24700392 2.130518e-13 TRUE  
## 4 FALSE 6.27608949 1.735972e-10 TRUE  
## 5 FALSE 5.93931038 1.431118e-09 TRUE  
## 6 FALSE 5.12440561 1.492389e-07 TRUE  
## 7 FALSE 5.12440561 1.492389e-07 TRUE  
## 8 FALSE 5.12440561 1.492389e-07 TRUE  
## 9 FALSE 5.12440561 1.492389e-07 TRUE  
## 10 FALSE 5.12440561 1.492389e-07 TRUE  
## 11 FALSE 5.12440561 1.492389e-07 TRUE  
## 12 FALSE 5.12440561 1.492389e-07 TRUE  
## 13 FALSE 5.12440561 1.492389e-07 TRUE  
## 14 FALSE 5.12440561 1.492389e-07 TRUE  
## 15 FALSE 5.12440561 1.492389e-07 TRUE  
## 16 FALSE 1.58376653 5.662343e-02 FALSE  
## 17 FALSE 0.76198509 2.230345e-01 FALSE  
## 18 FALSE 0.50018637 3.084719e-01 FALSE  
## 19 FALSE 0.50018637 3.084719e-01 FALSE  
## 20 FALSE 0.49065814 3.118341e-01 FALSE  
## 21 FALSE 0.49065814 3.118341e-01 FALSE  
## 22 FALSE 0.43993230 3.299931e-01 FALSE  
## 23 FALSE 0.39468524 3.465376e-01 FALSE  
## 24 FALSE 0.32950184 3.708882e-01 FALSE  
## 25 FALSE 0.32950184 3.708882e-01 FALSE  
## 26 FALSE 0.31484991 3.764378e-01 FALSE  
## 27 FALSE 0.28328161 3.884805e-01 FALSE  
## 28 FALSE 0.28328161 3.884805e-01 FALSE  
## 29 FALSE 0.27482353 3.917259e-01 FALSE  
## 30 FALSE 0.26609674 3.950824e-01 FALSE  
## 31 FALSE 0.26609674 3.950824e-01 FALSE  
## 32 FALSE 0.25707387 3.985609e-01 FALSE  
## 33 FALSE 0.25707387 3.985609e-01 FALSE  
## 34 FALSE 0.24772258 4.021745e-01 FALSE  
## 35 FALSE 0.23800416 4.059389e-01 FALSE  
## 36 FALSE 0.22787163 4.098730e-01 FALSE  
## 37 FALSE 0.22787163 4.098730e-01 FALSE  
## 38 FALSE 0.21726708 4.140001e-01 FALSE  
## 39 FALSE 0.21726708 4.140001e-01 FALSE  
## 40 FALSE 0.21726708 4.140001e-01 FALSE  
## 41 FALSE 0.20611765 4.183495e-01 FALSE  
## 42 FALSE 0.20611765 4.183495e-01 FALSE  
## 43 FALSE 0.19432958 4.229589e-01 FALSE  
## 44 FALSE 0.19432958 4.229589e-01 FALSE  
## 45 FALSE 0.19432958 4.229589e-01 FALSE  
## 46 FALSE 0.19432958 4.229589e-01 FALSE  
## 47 FALSE 0.19432958 4.229589e-01 FALSE  
## 48 FALSE 0.19432958 4.229589e-01 FALSE  
## 49 FALSE 0.18177868 4.278782e-01 FALSE  
## 50 FALSE 0.18177868 4.278782e-01 FALSE  
## 51 FALSE 0.18177868 4.278782e-01 FALSE  
## 52 FALSE 0.18177868 4.278782e-01 FALSE  
## 53 FALSE 0.18177868 4.278782e-01 FALSE  
## 54 FALSE 0.18177868 4.278782e-01 FALSE  
## 55 FALSE 0.18177868 4.278782e-01 FALSE  
## 56 FALSE 0.18177868 4.278782e-01 FALSE  
## 57 FALSE 0.18177868 4.278782e-01 FALSE  
## 58 FALSE 0.18177868 4.278782e-01 FALSE  
## 59 FALSE 0.18177868 4.278782e-01 FALSE  
## 60 FALSE 0.16829435 4.331759e-01 FALSE  
## 61 FALSE 0.16829435 4.331759e-01 FALSE  
## 62 FALSE 0.16829435 4.331759e-01 FALSE  
## 63 FALSE 0.16829435 4.331759e-01 FALSE  
## 64 FALSE 0.16829435 4.331759e-01 FALSE  
## 65 FALSE 0.16829435 4.331759e-01 FALSE  
## 66 FALSE 0.16829435 4.331759e-01 FALSE  
## 67 FALSE 0.15363102 4.389503e-01 FALSE  
## 68 FALSE 0.15363102 4.389503e-01 FALSE  
## 69 FALSE 0.15363102 4.389503e-01 FALSE  
## 70 FALSE 0.15363102 4.389503e-01 FALSE  
## 71 FALSE 0.15363102 4.389503e-01 FALSE  
## 72 FALSE 0.15363102 4.389503e-01 FALSE  
## 73 FALSE 0.15363102 4.389503e-01 FALSE  
## 74 FALSE 0.15363102 4.389503e-01 FALSE  
## 75 FALSE 0.15363102 4.389503e-01 FALSE  
## 76 FALSE 0.13741176 4.453527e-01 FALSE  
## 77 FALSE 0.13741176 4.453527e-01 FALSE  
## 78 FALSE 0.13741176 4.453527e-01 FALSE  
## 79 FALSE 0.13741176 4.453527e-01 FALSE  
## 80 FALSE 0.13741176 4.453527e-01 FALSE  
## 81 FALSE 0.13741176 4.453527e-01 FALSE  
## 82 FALSE 0.13741176 4.453527e-01 FALSE  
## 83 FALSE 0.13741176 4.453527e-01 FALSE  
## 84 FALSE 0.11900208 4.526369e-01 FALSE  
## 85 FALSE 0.11900208 4.526369e-01 FALSE  
## 86 FALSE 0.11900208 4.526369e-01 FALSE  
## 87 FALSE 0.11900208 4.526369e-01 FALSE  
## 88 FALSE 0.11900208 4.526369e-01 FALSE  
## 89 FALSE 0.11900208 4.526369e-01 FALSE  
## 90 FALSE 0.11900208 4.526369e-01 FALSE  
## 91 FALSE 0.11900208 4.526369e-01 FALSE  
## 92 FALSE 0.11900208 4.526369e-01 FALSE  
## 93 FALSE 0.11900208 4.526369e-01 FALSE  
## 94 FALSE 0.11900208 4.526369e-01 FALSE  
## 95 FALSE 0.11900208 4.526369e-01 FALSE  
## 96 FALSE 0.11900208 4.526369e-01 FALSE  
## 97 FALSE 0.11900208 4.526369e-01 FALSE  
## 98 FALSE 0.11900208 4.526369e-01 FALSE  
## 99 FALSE 0.11900208 4.526369e-01 FALSE  
## 100 FALSE 0.11900208 4.526369e-01 FALSE  
## 101 FALSE 0.11900208 4.526369e-01 FALSE  
## 102 FALSE 0.11900208 4.526369e-01 FALSE  
## 103 FALSE 0.11900208 4.526369e-01 FALSE  
## 104 FALSE 0.11900208 4.526369e-01 FALSE  
## 105 FALSE 0.11900208 4.526369e-01 FALSE  
## 106 FALSE 0.11900208 4.526369e-01 FALSE  
## 107 FALSE 0.11900208 4.526369e-01 FALSE  
## 108 FALSE 0.11900208 4.526369e-01 FALSE  
## 109 FALSE 0.11900208 4.526369e-01 FALSE  
## 110 FALSE 0.11900208 4.526369e-01 FALSE  
## 111 FALSE 0.11900208 4.526369e-01 FALSE  
## 112 FALSE 0.11900208 4.526369e-01 FALSE  
## 113 FALSE 0.11900208 4.526369e-01 FALSE  
## 114 FALSE 0.11900208 4.526369e-01 FALSE  
## 115 FALSE 0.11900208 4.526369e-01 FALSE  
## 116 FALSE 0.11900208 4.526369e-01 FALSE  
## 117 FALSE 0.11900208 4.526369e-01 FALSE  
## 118 FALSE 0.09716479 4.612978e-01 FALSE  
## 119 FALSE 0.09716479 4.612978e-01 FALSE  
## 120 FALSE 0.09716479 4.612978e-01 FALSE  
## 121 FALSE 0.09716479 4.612978e-01 FALSE  
## 122 FALSE 0.09716479 4.612978e-01 FALSE  
## 123 FALSE 0.09716479 4.612978e-01 FALSE  
## 124 FALSE 0.09716479 4.612978e-01 FALSE  
## 125 FALSE 0.09716479 4.612978e-01 FALSE  
## 126 FALSE 0.09716479 4.612978e-01 FALSE  
## 127 FALSE 0.09716479 4.612978e-01 FALSE  
## 128 FALSE 0.09716479 4.612978e-01 FALSE  
## 129 FALSE 0.09716479 4.612978e-01 FALSE  
## 130 FALSE 0.09716479 4.612978e-01 FALSE  
## 131 FALSE 0.09716479 4.612978e-01 FALSE  
## 132 FALSE 0.09716479 4.612978e-01 FALSE  
## 133 FALSE 0.09716479 4.612978e-01 FALSE  
## 134 FALSE 0.09716479 4.612978e-01 FALSE  
## 135 FALSE 0.09716479 4.612978e-01 FALSE  
## 136 FALSE 0.09716479 4.612978e-01 FALSE  
## 137 FALSE 0.09716479 4.612978e-01 FALSE  
## 138 FALSE 0.09716479 4.612978e-01 FALSE  
## 139 FALSE 0.09716479 4.612978e-01 FALSE  
## 140 FALSE 0.09716479 4.612978e-01 FALSE  
## 141 FALSE 0.09716479 4.612978e-01 FALSE  
## 142 FALSE 0.09716479 4.612978e-01 FALSE  
## 143 FALSE 0.09716479 4.612978e-01 FALSE  
## 144 FALSE 0.09716479 4.612978e-01 FALSE  
## 145 FALSE 0.09716479 4.612978e-01 FALSE  
## 146 FALSE 0.09716479 4.612978e-01 FALSE  
## 147 FALSE 0.09716479 4.612978e-01 FALSE  
## 148 FALSE 0.09716479 4.612978e-01 FALSE  
## 149 FALSE 0.09716479 4.612978e-01 FALSE  
## 150 FALSE 0.09716479 4.612978e-01 FALSE  
## 151 FALSE 0.09716479 4.612978e-01 FALSE  
## 152 FALSE 0.09716479 4.612978e-01 FALSE  
## 153 FALSE 0.09716479 4.612978e-01 FALSE  
## 154 FALSE 0.09716479 4.612978e-01 FALSE  
## 155 FALSE 0.09716479 4.612978e-01 FALSE  
## 156 FALSE 0.09716479 4.612978e-01 FALSE  
## 157 FALSE 0.09716479 4.612978e-01 FALSE  
## 158 FALSE 0.09716479 4.612978e-01 FALSE  
## 159 FALSE 0.09716479 4.612978e-01 FALSE  
## 160 FALSE 0.09716479 4.612978e-01 FALSE  
## 161 FALSE 0.09716479 4.612978e-01 FALSE  
## 162 FALSE 0.09716479 4.612978e-01 FALSE  
## 163 FALSE 0.09716479 4.612978e-01 FALSE  
## 164 FALSE 0.09716479 4.612978e-01 FALSE  
## 165 FALSE 0.09716479 4.612978e-01 FALSE  
## 166 FALSE 0.09716479 4.612978e-01 FALSE  
## 167 FALSE 0.09716479 4.612978e-01 FALSE  
## 168 FALSE 0.09716479 4.612978e-01 FALSE  
## 169 FALSE 0.09716479 4.612978e-01 FALSE  
## 170 FALSE 0.09716479 4.612978e-01 FALSE  
## 171 FALSE 0.09716479 4.612978e-01 FALSE  
## 172 FALSE 0.09716479 4.612978e-01 FALSE  
## 173 FALSE 0.09716479 4.612978e-01 FALSE  
## 174 FALSE 0.09716479 4.612978e-01 FALSE  
## 175 FALSE 0.09716479 4.612978e-01 FALSE  
## 176 FALSE 0.09716479 4.612978e-01 FALSE  
## 177 FALSE 0.09716479 4.612978e-01 FALSE  
## 178 FALSE 0.09716479 4.612978e-01 FALSE  
## 179 FALSE 0.09716479 4.612978e-01 FALSE  
## 180 FALSE 0.09716479 4.612978e-01 FALSE  
## 181 FALSE 0.09716479 4.612978e-01 FALSE  
## 182 FALSE 0.09716479 4.612978e-01 FALSE  
## 183 FALSE 0.09716479 4.612978e-01 FALSE  
## 184 FALSE 0.09716479 4.612978e-01 FALSE  
## 185 FALSE 0.09716479 4.612978e-01 FALSE  
## 186 FALSE 0.09716479 4.612978e-01 FALSE  
## 187 FALSE 0.09716479 4.612978e-01 FALSE  
## 188 FALSE 0.09716479 4.612978e-01 FALSE  
## 189 FALSE 0.09716479 4.612978e-01 FALSE  
## 190 FALSE 0.09716479 4.612978e-01 FALSE  
## 191 FALSE 0.09716479 4.612978e-01 FALSE  
## 192 FALSE 0.09716479 4.612978e-01 FALSE  
## 193 FALSE 0.09716479 4.612978e-01 FALSE  
## 194 FALSE 0.09716479 4.612978e-01 FALSE  
## 195 FALSE 0.09716479 4.612978e-01 FALSE  
## 196 FALSE 0.09716479 4.612978e-01 FALSE  
## 197 FALSE 0.09716479 4.612978e-01 FALSE  
## 198 FALSE 0.09716479 4.612978e-01 FALSE  
## 199 FALSE 0.09716479 4.612978e-01 FALSE  
## 200 FALSE 0.09716479 4.612978e-01 FALSE  
## 201 FALSE 0.09716479 4.612978e-01 FALSE  
## 202 FALSE 0.09716479 4.612978e-01 FALSE  
## 203 FALSE 0.09716479 4.612978e-01 FALSE  
## 204 FALSE 0.09716479 4.612978e-01 FALSE  
## 205 FALSE 0.09716479 4.612978e-01 FALSE  
## 206 FALSE 0.09716479 4.612978e-01 FALSE  
## 207 FALSE 0.09716479 4.612978e-01 FALSE  
## 208 FALSE 0.09716479 4.612978e-01 FALSE  
## 209 FALSE 0.09716479 4.612978e-01 FALSE  
## 210 FALSE 0.09716479 4.612978e-01 FALSE  
## 211 FALSE 0.09716479 4.612978e-01 FALSE  
## 212 FALSE 0.09716479 4.612978e-01 FALSE  
## 213 FALSE 0.09716479 4.612978e-01 FALSE  
## 214 FALSE 0.09716479 4.612978e-01 FALSE  
## 215 FALSE 0.09716479 4.612978e-01 FALSE  
## 216 FALSE 0.09716479 4.612978e-01 FALSE  
##   
##   
## Summary statistics:  
##   
## Total Chi squared = 542.4691   
## Total degrees of freedom = 214   
## p = 0   
## Sum of counts = 2139   
##   
## Levels:  
##   
## Data X   
## 2 215

r7 <- cfa::cfa(configs, counts)  
# save to disc  
sig\_words <- r7$table %>%  
 as.data.frame() %>%  
 dplyr::mutate(expected = round(expected, 1),  
 Q = round(Q, 3),  
 chisq = round(chisq, 3),  
 z = round(z, 3),  
 p.z = round(p.z, 3),  
 label = stringr::str\_remove\_all(label, "Xmuch ")) %>%  
 dplyr::select(-p.chisq, -sig.chisq)  
# save to disc  
write.table(sig\_words,   
 here::here("tables", "xmuch\_words.txt"),   
 sep = "\t",   
 row.names = F)  
# extract words  
words <- r7$table %>%  
 as.data.frame() %>%  
 dplyr::mutate(Type = ifelse(n > expected, "Type", "Antitype")) %>%  
 dplyr::filter(sig.z == T,  
 Type == "Type") %>%  
 dplyr::pull(label) %>%  
 stringr::str\_remove\_all("Xmuch ")  
words

## [1] "jealous" "paranoid" "agenda" "threatened" "other"   
## [6] "hypocrite" "hipster" "generalise" "excuses" "excited"   
## [11] "entitled" "creepy" "coincidence" "bitter" "biased"

Words that are significantly over-represented in the X-much structure: jealous, paranoid, agenda, threatened, other, hypocrite, hipster, generalise, excuses, excited, entitled, creepy, coincidence, bitter, biased

### X2 (w variety)

RQ: Do the words that are significantly attracted by the x-much construction differ across varieties?

wordcat\_tb <- xmuch %>%  
 dplyr::filter(Data == "Xmuch") %>%  
 dplyr::group\_by(Variety, X) %>%  
 dplyr::mutate(Frequency = n()) %>%  
 dplyr::mutate(X = ifelse(Frequency < 2, "other", X)) %>%  
 dplyr::ungroup() %>%  
 dplyr::group\_by(X, Variety) %>%  
 dplyr::summarise(Frequency = n())

## `summarise()` has grouped output by 'X'. You can override using the `.groups` argument.

configs <- wordcat\_tb %>%  
 dplyr::select(Variety, X)  
counts = wordcat\_tb$Frequency  
cfa::cfa(configs, counts)

##   
## \*\*\* Analysis of configuration frequencies (CFA) \*\*\*  
##   
## label n expected Q chisq p.chisq  
## 1 Ireland excuses 2 0.4647887 0.010846850 5.070849338 0.8863952  
## 2 Ireland coincidence 2 0.4647887 0.010846850 5.070849338 0.8863952  
## 3 Ireland agenda 2 0.9295775 0.007587859 1.232607768 0.9995549  
## 4 Australia jealous 7 5.3732394 0.011906603 0.492505492 0.9999939  
## 5 Australia agenda 2 3.0704225 0.007704785 0.373174829 0.9999984  
## 6 Australia threatened 3 2.3028169 0.004990674 0.211073782 0.9999999  
## 7 Australia paranoid 3 2.3028169 0.004990674 0.211073782 0.9999999  
## 8 Australia hypocrite 2 1.5352113 0.003308934 0.140715855 1.0000000  
## 9 Australia generalise 2 1.5352113 0.003308934 0.140715855 1.0000000  
## 10 Australia entitled 2 1.5352113 0.003308934 0.140715855 1.0000000  
## 11 Australia creepy 2 1.5352113 0.003308934 0.140715855 1.0000000  
## 12 Ireland other 27 26.2605634 0.006388804 0.020820822 1.0000000  
## 13 Australia other 86 86.7394366 0.008524803 0.006303552 1.0000000  
## sig.chisq z p.z sig.z  
## 1 FALSE 2.25185464 0.01216573 FALSE  
## 2 FALSE 2.25185464 0.01216573 FALSE  
## 3 FALSE 1.11022870 0.13345024 FALSE  
## 4 FALSE 0.70178736 0.24140589 FALSE  
## 5 FALSE 0.61088037 0.27063939 FALSE  
## 6 FALSE 0.45942767 0.32296354 FALSE  
## 7 FALSE 0.45942767 0.32296354 FALSE  
## 8 FALSE 0.37512112 0.35378520 FALSE  
## 9 FALSE 0.37512112 0.35378520 FALSE  
## 10 FALSE 0.37512112 0.35378520 FALSE  
## 11 FALSE 0.37512112 0.35378520 FALSE  
## 12 FALSE 0.14429422 0.44263407 FALSE  
## 13 FALSE 0.07939491 0.46835926 FALSE  
##   
##   
## Summary statistics:  
##   
## Total Chi squared = 13.25212   
## Total degrees of freedom = 10   
## p = 0.209915   
## Sum of counts = 142   
##   
## Levels:  
##   
## Variety X   
## 2 11

r8 <- cfa::cfa(configs, counts)  
words <- r8$table %>%  
 as.data.frame() %>%  
 dplyr::filter(sig.z == T) %>%  
 dplyr::pull(label) %>%  
 stringr::str\_remove\_all("Xmuch ")  
words

## character(0)

The analysis does not confirm variety specific attraction of types to the X-much structure.

### CFA (w variety)

RQ: Do the words that are significantly attracted by the x-much construction differ across varieties when we consider the control data?

wordcat\_tb <- xmuch %>%  
 dplyr::group\_by(Data, Variety, X) %>%  
 dplyr::mutate(Frequency = n()) %>%  
 dplyr::mutate(X = ifelse(Frequency < 2, "other", X)) %>%  
 dplyr::ungroup() %>%  
 dplyr::group\_by(X, Variety, Status) %>%  
 dplyr::summarise(Frequency = n())

## `summarise()` has grouped output by 'X', 'Variety'. You can override using the `.groups` argument.

configs <- wordcat\_tb %>%  
 dplyr::select(Status, Variety, X)  
counts = wordcat\_tb$Frequency  
cfa::cfa(configs, counts)

##   
## \*\*\* Analysis of configuration frequencies (CFA) \*\*\*  
##   
## label n expected Q chisq p.chisq  
## 1 phrase Australia other 31 9.9007802 9.905276e-03 4.496384e+01 1  
## 2 word Ireland no 7 3.3142346 1.724992e-03 4.098945e+00 1  
## 3 word Ireland documentary 7 3.3142346 1.724992e-03 4.098945e+00 1  
## 4 word Australia jealous 7 3.5516532 1.614056e-03 3.348045e+00 1  
## 5 word Ireland on 13 8.0488556 2.322354e-03 3.045629e+00 1  
## 6 word Australia like 6 3.0442742 1.383148e-03 2.869753e+00 1  
## 7 word Australia if 6 3.0442742 1.383148e-03 2.869753e+00 1  
## 8 word Ireland the 71 58.2358374 6.131416e-03 2.797656e+00 1  
## 9 word Australia on 4 8.6254435 2.170169e-03 2.480421e+00 1  
## 10 word Ireland get 4 1.8938484 9.850548e-04 2.342254e+00 1  
## 11 word Ireland of 26 19.4119458 3.106711e-03 2.235863e+00 1  
## 12 word Australia he 9 5.5811693 1.601762e-03 2.094257e+00 1  
## 13 word Ireland he 2 5.2080830 1.502761e-03 1.976120e+00 1  
## 14 word Australia think 4 2.0295161 9.216609e-04 1.913169e+00 1  
## 15 word Australia community 4 2.0295161 9.216609e-04 1.913169e+00 1  
## 16 word Ireland a 32 25.0934909 3.265633e-03 1.900886e+00 1  
## 17 word Ireland rt 3 1.4203863 7.386275e-04 1.756691e+00 1  
## 18 word Ireland part 3 1.4203863 7.386275e-04 1.756691e+00 1  
## 19 word Ireland new 3 1.4203863 7.386275e-04 1.756691e+00 1  
## 20 word Ireland bored 3 1.4203863 7.386275e-04 1.756691e+00 1  
## 21 word Ireland before 3 1.4203863 7.386275e-04 1.756691e+00 1  
## 22 word Ireland 2011 3 1.4203863 7.386275e-04 1.756691e+00 1  
## 23 word Australia the 52 62.4076205 5.009462e-03 1.735662e+00 1  
## 24 word Australia they 8 5.0737903 1.370637e-03 1.687634e+00 1  
## 25 word Australia of 15 20.8025402 2.738084e-03 1.618527e+00 1  
## 26 word Ireland they 2 4.7346209 1.280694e-03 1.579462e+00 1  
## 27 word Australia very 3 1.5221371 6.910817e-04 1.434877e+00 1  
## 28 word Australia us 3 1.5221371 6.910817e-04 1.434877e+00 1  
## 29 word Australia threatened 3 1.5221371 6.910817e-04 1.434877e+00 1  
## 30 word Australia services 3 1.5221371 6.910817e-04 1.434877e+00 1  
## 31 word Australia said 3 1.5221371 6.910817e-04 1.434877e+00 1  
## 32 word Australia paranoid 3 1.5221371 6.910817e-04 1.434877e+00 1  
## 33 word Australia lead 3 1.5221371 6.910817e-04 1.434877e+00 1  
## 34 word Australia had 3 1.5221371 6.910817e-04 1.434877e+00 1  
## 35 word Australia government 3 1.5221371 6.910817e-04 1.434877e+00 1  
## 36 word Australia being 3 1.5221371 6.910817e-04 1.434877e+00 1  
## 37 word Ireland other 447 472.9886307 1.558995e-02 1.427960e+00 1  
## 38 word Ireland have 6 3.7876967 1.035620e-03 1.292154e+00 1  
## 39 word Australia a 21 26.8910885 2.787877e-03 1.290573e+00 1  
## 40 word Ireland years 2 0.9469242 4.923093e-04 1.171127e+00 1  
## 41 word Ireland world 2 0.9469242 4.923093e-04 1.171127e+00 1  
## 42 word Ireland work 2 0.9469242 4.923093e-04 1.171127e+00 1  
## 43 word Ireland within 2 0.9469242 4.923093e-04 1.171127e+00 1  
## 44 word Ireland well 2 0.9469242 4.923093e-04 1.171127e+00 1  
## 45 word Ireland vulnerable 2 0.9469242 4.923093e-04 1.171127e+00 1  
## 46 word Ireland two 2 0.9469242 4.923093e-04 1.171127e+00 1  
## 47 word Ireland town 2 0.9469242 4.923093e-04 1.171127e+00 1  
## 48 word Ireland total 2 0.9469242 4.923093e-04 1.171127e+00 1  
## 49 word Ireland too 2 0.9469242 4.923093e-04 1.171127e+00 1  
## 50 word Ireland though 2 0.9469242 4.923093e-04 1.171127e+00 1  
## 51 word Ireland such 2 0.9469242 4.923093e-04 1.171127e+00 1  
## 52 word Ireland stay 2 0.9469242 4.923093e-04 1.171127e+00 1  
## 53 word Ireland state 2 0.9469242 4.923093e-04 1.171127e+00 1  
## 54 word Ireland put 2 0.9469242 4.923093e-04 1.171127e+00 1  
## 55 word Ireland over 2 0.9469242 4.923093e-04 1.171127e+00 1  
## 56 word Ireland one 2 0.9469242 4.923093e-04 1.171127e+00 1  
## 57 word Ireland number 2 0.9469242 4.923093e-04 1.171127e+00 1  
## 58 word Ireland more 2 0.9469242 4.923093e-04 1.171127e+00 1  
## 59 word Ireland life 2 0.9469242 4.923093e-04 1.171127e+00 1  
## 60 word Ireland just 2 0.9469242 4.923093e-04 1.171127e+00 1  
## 61 word Ireland ireland 2 0.9469242 4.923093e-04 1.171127e+00 1  
## 62 word Ireland http 2 0.9469242 4.923093e-04 1.171127e+00 1  
## 63 word Ireland home 2 0.9469242 4.923093e-04 1.171127e+00 1  
## 64 word Ireland her 2 0.9469242 4.923093e-04 1.171127e+00 1  
## 65 word Ireland god 2 0.9469242 4.923093e-04 1.171127e+00 1  
## 66 word Ireland four 2 0.9469242 4.923093e-04 1.171127e+00 1  
## 67 word Ireland find 2 0.9469242 4.923093e-04 1.171127e+00 1  
## 68 word Ireland fact 2 0.9469242 4.923093e-04 1.171127e+00 1  
## 69 word Ireland excuses 2 0.9469242 4.923093e-04 1.171127e+00 1  
## 70 word Ireland early 2 0.9469242 4.923093e-04 1.171127e+00 1  
## 71 word Ireland ear 2 0.9469242 4.923093e-04 1.171127e+00 1  
## 72 word Ireland different 2 0.9469242 4.923093e-04 1.171127e+00 1  
## 73 word Ireland didn 2 0.9469242 4.923093e-04 1.171127e+00 1  
## 74 word Ireland did 2 0.9469242 4.923093e-04 1.171127e+00 1  
## 75 word Ireland day 2 0.9469242 4.923093e-04 1.171127e+00 1  
## 76 word Ireland cork 2 0.9469242 4.923093e-04 1.171127e+00 1  
## 77 word Ireland computer 2 0.9469242 4.923093e-04 1.171127e+00 1  
## 78 word Ireland coincidence 2 0.9469242 4.923093e-04 1.171127e+00 1  
## 79 word Ireland brand 2 0.9469242 4.923093e-04 1.171127e+00 1  
## 80 word Ireland bill 2 0.9469242 4.923093e-04 1.171127e+00 1  
## 81 word Ireland begin 2 0.9469242 4.923093e-04 1.171127e+00 1  
## 82 word Ireland been 2 0.9469242 4.923093e-04 1.171127e+00 1  
## 83 word Ireland bear 2 0.9469242 4.923093e-04 1.171127e+00 1  
## 84 word Ireland another 2 0.9469242 4.923093e-04 1.171127e+00 1  
## 85 word Ireland all 2 0.9469242 4.923093e-04 1.171127e+00 1  
## 86 word Ireland aid 2 0.9469242 4.923093e-04 1.171127e+00 1  
## 87 word Ireland after 2 0.9469242 4.923093e-04 1.171127e+00 1  
## 88 word Australia have 2 4.0590322 9.639930e-04 1.044489e+00 1  
## 89 word Australia and 32 26.8910885 2.417723e-03 9.706181e-01 1  
## 90 word Australia will 2 1.0147581 4.606118e-04 9.565844e-01 1  
## 91 word Australia were 2 1.0147581 4.606118e-04 9.565844e-01 1  
## 92 word Australia ways 2 1.0147581 4.606118e-04 9.565844e-01 1  
## 93 word Australia unless 2 1.0147581 4.606118e-04 9.565844e-01 1  
## 94 word Australia types 2 1.0147581 4.606118e-04 9.565844e-01 1  
## 95 word Australia turn 2 1.0147581 4.606118e-04 9.565844e-01 1  
## 96 word Australia these 2 1.0147581 4.606118e-04 9.565844e-01 1  
## 97 word Australia thanks 2 1.0147581 4.606118e-04 9.565844e-01 1  
## 98 word Australia system 2 1.0147581 4.606118e-04 9.565844e-01 1  
## 99 word Australia school 2 1.0147581 4.606118e-04 9.565844e-01 1  
## 100 word Australia played 2 1.0147581 4.606118e-04 9.565844e-01 1  
## 101 word Australia people 2 1.0147581 4.606118e-04 9.565844e-01 1  
## 102 word Australia paragraph 2 1.0147581 4.606118e-04 9.565844e-01 1  
## 103 word Australia most 2 1.0147581 4.606118e-04 9.565844e-01 1  
## 104 word Australia months 2 1.0147581 4.606118e-04 9.565844e-01 1  
## 105 word Australia making 2 1.0147581 4.606118e-04 9.565844e-01 1  
## 106 word Australia made 2 1.0147581 4.606118e-04 9.565844e-01 1  
## 107 word Australia location 2 1.0147581 4.606118e-04 9.565844e-01 1  
## 108 word Australia known 2 1.0147581 4.606118e-04 9.565844e-01 1  
## 109 word Australia know 2 1.0147581 4.606118e-04 9.565844e-01 1  
## 110 word Australia into 2 1.0147581 4.606118e-04 9.565844e-01 1  
## 111 word Australia impressive 2 1.0147581 4.606118e-04 9.565844e-01 1  
## 112 word Australia important 2 1.0147581 4.606118e-04 9.565844e-01 1  
## 113 word Australia hypocrite 2 1.0147581 4.606118e-04 9.565844e-01 1  
## 114 word Australia how 2 1.0147581 4.606118e-04 9.565844e-01 1  
## 115 word Australia hours 2 1.0147581 4.606118e-04 9.565844e-01 1  
## 116 word Australia hair 2 1.0147581 4.606118e-04 9.565844e-01 1  
## 117 word Australia good 2 1.0147581 4.606118e-04 9.565844e-01 1  
## 118 word Australia generalise 2 1.0147581 4.606118e-04 9.565844e-01 1  
## 119 word Australia further 2 1.0147581 4.606118e-04 9.565844e-01 1  
## 120 word Australia found 2 1.0147581 4.606118e-04 9.565844e-01 1  
## 121 word Australia every 2 1.0147581 4.606118e-04 9.565844e-01 1  
## 122 word Australia entitled 2 1.0147581 4.606118e-04 9.565844e-01 1  
## 123 word Australia else 2 1.0147581 4.606118e-04 9.565844e-01 1  
## 124 word Australia down 2 1.0147581 4.606118e-04 9.565844e-01 1  
## 125 word Australia creepy 2 1.0147581 4.606118e-04 9.565844e-01 1  
## 126 word Australia church 2 1.0147581 4.606118e-04 9.565844e-01 1  
## 127 word Australia business 2 1.0147581 4.606118e-04 9.565844e-01 1  
## 128 word Australia because 2 1.0147581 4.606118e-04 9.565844e-01 1  
## 129 word Australia beautiful 2 1.0147581 4.606118e-04 9.565844e-01 1  
## 130 word Australia ball 2 1.0147581 4.606118e-04 9.565844e-01 1  
## 131 word Australia back 2 1.0147581 4.606118e-04 9.565844e-01 1  
## 132 word Australia australia 2 1.0147581 4.606118e-04 9.565844e-01 1  
## 133 word Australia also 2 1.0147581 4.606118e-04 9.565844e-01 1  
## 134 word Australia act 2 1.0147581 4.606118e-04 9.565844e-01 1  
## 135 word Ireland was 13 9.9427039 1.435312e-03 9.400923e-01 1  
## 136 word Ireland for 14 10.8896281 1.460879e-03 8.884062e-01 1  
## 137 word Ireland can 5 3.3142346 7.889627e-04 8.574543e-01 1  
## 138 word Australia that 10 7.6106854 1.120487e-03 7.501064e-01 1  
## 139 word Ireland at 6 4.2611588 8.141638e-04 7.095649e-01 1  
## 140 word Australia can 2 3.5516532 7.262770e-04 6.778893e-01 1  
## 141 word Ireland and 21 25.0934909 1.935542e-03 6.677695e-01 1  
## 142 word Australia was 8 10.6549596 1.246843e-03 6.615521e-01 1  
## 143 word Ireland that 5 7.1019314 9.854814e-04 6.221006e-01 1  
## 144 word Ireland in 28 24.1465667 1.821219e-03 6.149507e-01 1  
## 145 word Australia for 9 11.6697177 1.254372e-03 6.107596e-01 1  
## 146 word Australia their 5 3.5516532 6.779227e-04 5.906287e-01 1  
## 147 word Australia some 5 3.5516532 6.779227e-04 5.906287e-01 1  
## 148 word Australia out 5 3.5516532 6.779227e-04 5.906287e-01 1  
## 149 word Australia about 5 3.5516532 6.779227e-04 5.906287e-01 1  
## 150 word Australia at 3 4.5664113 7.335331e-04 5.373244e-01 1  
## 151 word Ireland their 2 3.3142346 6.150809e-04 5.211498e-01 1  
## 152 word Ireland some 2 3.3142346 6.150809e-04 5.211498e-01 1  
## 153 word Ireland out 2 3.3142346 6.150809e-04 5.211498e-01 1  
## 154 word Ireland about 2 3.3142346 6.150809e-04 5.211498e-01 1  
## 155 word Ireland who 4 2.8407726 5.424151e-04 4.730432e-01 1  
## 156 word Ireland are 10 8.0488556 9.151919e-04 4.729821e-01 1  
## 157 word Australia you 10 8.1180645 8.827579e-04 4.362717e-01 1  
## 158 word Ireland or 5 3.7876967 5.675013e-04 3.880140e-01 1  
## 159 word Ireland is 18 15.6242491 1.118329e-03 3.612457e-01 1  
## 160 word Australia who 2 3.0442742 4.886738e-04 3.582163e-01 1  
## 161 word Ireland there 6 4.7346209 5.926097e-04 3.381864e-01 1  
## 162 word Ireland you 6 7.5753935 7.387804e-04 3.276219e-01 1  
## 163 word Australia in 23 25.8763305 1.360531e-03 3.197237e-01 1  
## 164 word Australia are 7 8.6254435 7.626269e-04 3.063108e-01 1  
## 165 word Ireland from 7 5.6815451 6.177405e-04 3.059596e-01 1  
## 166 word Australia by 4 3.0442742 4.472371e-04 3.000426e-01 1  
## 167 word Australia or 3 4.0590322 4.958153e-04 2.763095e-01 1  
## 168 word Ireland by 2 2.8407726 3.934066e-04 2.488402e-01 1  
## 169 word Australia there 4 5.0737903 5.029637e-04 2.272513e-01 1  
## 170 word Australia which 5 4.0590322 4.405402e-04 2.181358e-01 1  
## 171 word Australia do 5 4.0590322 4.405402e-04 2.181358e-01 1  
## 172 word Australia from 5 6.0885483 5.101188e-04 1.946174e-01 1  
## 173 word Australia is 15 16.7435079 8.211481e-04 1.815522e-01 1  
## 174 word Australia to 28 25.8763305 1.004515e-03 1.742895e-01 1  
## 175 word Ireland we 3 2.3673105 2.959767e-04 1.690932e-01 1  
## 176 word Ireland which 3 3.7876967 3.687352e-04 1.638109e-01 1  
## 177 word Ireland do 3 3.7876967 3.687352e-04 1.638109e-01 1  
## 178 word Australia i 13 11.6697177 6.250357e-04 1.516447e-01 1  
## 179 word Ireland would 4 3.3142346 3.209482e-04 1.418952e-01 1  
## 180 word Ireland up 4 3.3142346 3.209482e-04 1.418952e-01 1  
## 181 word Ireland my 4 3.3142346 3.209482e-04 1.418952e-01 1  
## 182 word Ireland but 4 3.3142346 3.209482e-04 1.418952e-01 1  
## 183 word Ireland with 7 6.1550072 3.959954e-04 1.160052e-01 1  
## 184 word Australia we 2 2.5368951 2.511834e-04 1.136257e-01 1  
## 185 word Australia would 3 3.5516532 2.582104e-04 8.568439e-02 1  
## 186 word Australia up 3 3.5516532 2.582104e-04 8.568439e-02 1  
## 187 word Australia my 3 3.5516532 2.582104e-04 8.568439e-02 1  
## 188 word Australia but 3 3.5516532 2.582104e-04 8.568439e-02 1  
## 189 word Australia your 3 2.5368951 2.166610e-04 8.453881e-02 1  
## 190 word Australia way 3 2.5368951 2.166610e-04 8.453881e-02 1  
## 191 word Australia still 3 2.5368951 2.166610e-04 8.453881e-02 1  
## 192 word Australia our 3 2.5368951 2.166610e-04 8.453881e-02 1  
## 193 word Australia much 3 2.5368951 2.166610e-04 8.453881e-02 1  
## 194 word Australia his 3 2.5368951 2.166610e-04 8.453881e-02 1  
## 195 word Australia any 3 2.5368951 2.166610e-04 8.453881e-02 1  
## 196 word Ireland i 10 10.8896281 4.178403e-04 7.267817e-02 1  
## 197 phrase Ireland other 10 9.2389394 3.571778e-04 6.269261e-02 1  
## 198 word Ireland your 2 2.3673105 1.718305e-04 5.699167e-02 1  
## 199 word Ireland way 2 2.3673105 1.718305e-04 5.699167e-02 1  
## 200 word Ireland still 2 2.3673105 1.718305e-04 5.699167e-02 1  
## 201 word Ireland our 2 2.3673105 1.718305e-04 5.699167e-02 1  
## 202 word Ireland much 2 2.3673105 1.718305e-04 5.699167e-02 1  
## 203 word Ireland his 2 2.3673105 1.718305e-04 5.699167e-02 1  
## 204 word Ireland any 2 2.3673105 1.718305e-04 5.699167e-02 1  
## 205 word Ireland to 23 24.1465667 5.418933e-04 5.444315e-02 1  
## 206 word Australia with 6 6.5959274 2.793317e-04 5.384071e-02 1  
## 207 word Australia an 5 4.5664113 2.030448e-04 4.117001e-02 1  
## 208 word Australia other 511 506.8716497 2.527879e-03 3.362444e-02 1  
## 209 word Australia be 6 5.5811693 1.962270e-04 3.143054e-02 1  
## 210 word Ireland not 7 6.6284693 1.741519e-04 2.082458e-02 1  
## 211 word Ireland it 7 6.6284693 1.741519e-04 2.082458e-02 1  
## 212 word Australia as 8 7.6106854 1.825720e-04 1.991487e-02 1  
## 213 word Ireland an 4 4.2611588 1.222803e-04 1.600596e-02 1  
## 214 word Ireland this 5 4.7346209 1.242839e-04 1.487470e-02 1  
## 215 word Ireland so 4 3.7876967 9.938303e-05 1.189976e-02 1  
## 216 word Ireland what 3 2.8407726 7.450425e-05 8.924818e-03 1  
## 217 word Ireland time 3 2.8407726 7.450425e-05 8.924818e-03 1  
## 218 word Ireland may 3 2.8407726 7.450425e-05 8.924818e-03 1  
## 219 word Ireland has 3 2.8407726 7.450425e-05 8.924818e-03 1  
## 220 word Ireland be 5 5.2080830 9.747227e-05 8.313720e-03 1  
## 221 word Ireland want 2 1.8938484 4.964750e-05 5.949879e-03 1  
## 222 word Ireland agenda 2 1.8938484 4.964750e-05 5.949879e-03 1  
## 223 word Australia not 7 7.1033064 4.843479e-05 1.502429e-03 1  
## 224 word Australia it 7 7.1033064 4.843479e-05 1.502429e-03 1  
## 225 word Ireland as 7 7.1019314 4.779009e-05 1.462984e-03 1  
## 226 word Australia this 5 5.0737903 3.456339e-05 1.073163e-03 1  
## 227 word Australia so 4 4.0590322 2.763758e-05 8.585308e-04 1  
## 228 word Australia what 3 3.0442742 2.071834e-05 6.438981e-04 1  
## 229 word Australia time 3 3.0442742 2.071834e-05 6.438981e-04 1  
## 230 word Australia may 3 3.0442742 2.071834e-05 6.438981e-04 1  
## 231 word Australia has 3 3.0442742 2.071834e-05 6.438981e-04 1  
## 232 word Australia want 2 2.0295161 1.380567e-05 4.292654e-04 1  
## 233 word Australia agenda 2 2.0295161 1.380567e-05 4.292654e-04 1  
## sig.chisq z p.z sig.z  
## 1 FALSE 6.70550803 1.003531e-11 TRUE  
## 2 FALSE 2.02458519 2.145499e-02 FALSE  
## 3 FALSE 2.02458519 2.145499e-02 FALSE  
## 4 FALSE 1.82976645 3.364243e-02 FALSE  
## 5 FALSE 1.74517317 4.047737e-02 FALSE  
## 6 FALSE 1.69403456 4.512936e-02 FALSE  
## 7 FALSE 1.69403456 4.512936e-02 FALSE  
## 8 FALSE 1.67261954 4.720111e-02 FALSE  
## 9 FALSE 1.57493510 5.763571e-02 FALSE  
## 10 FALSE 1.53044255 6.295361e-02 FALSE  
## 11 FALSE 1.49528036 6.742064e-02 FALSE  
## 12 FALSE 1.44715482 7.392678e-02 FALSE  
## 13 FALSE 1.40574533 7.989988e-02 FALSE  
## 14 FALSE 1.38317342 8.330585e-02 FALSE  
## 15 FALSE 1.38317342 8.330585e-02 FALSE  
## 16 FALSE 1.37872625 8.398959e-02 FALSE  
## 17 FALSE 1.32540213 9.251891e-02 FALSE  
## 18 FALSE 1.32540213 9.251891e-02 FALSE  
## 19 FALSE 1.32540213 9.251891e-02 FALSE  
## 20 FALSE 1.32540213 9.251891e-02 FALSE  
## 21 FALSE 1.32540213 9.251891e-02 FALSE  
## 22 FALSE 1.32540213 9.251891e-02 FALSE  
## 23 FALSE 1.31744544 9.384468e-02 FALSE  
## 24 FALSE 1.29908985 9.695655e-02 FALSE  
## 25 FALSE 1.27221342 1.016487e-01 FALSE  
## 26 FALSE 1.25676630 1.044191e-01 FALSE  
## 27 FALSE 1.19786332 1.154851e-01 FALSE  
## 28 FALSE 1.19786332 1.154851e-01 FALSE  
## 29 FALSE 1.19786332 1.154851e-01 FALSE  
## 30 FALSE 1.19786332 1.154851e-01 FALSE  
## 31 FALSE 1.19786332 1.154851e-01 FALSE  
## 32 FALSE 1.19786332 1.154851e-01 FALSE  
## 33 FALSE 1.19786332 1.154851e-01 FALSE  
## 34 FALSE 1.19786332 1.154851e-01 FALSE  
## 35 FALSE 1.19786332 1.154851e-01 FALSE  
## 36 FALSE 1.19786332 1.154851e-01 FALSE  
## 37 FALSE 1.19497287 1.160488e-01 FALSE  
## 38 FALSE 1.13672931 1.278257e-01 FALSE  
## 39 FALSE 1.13603407 1.279712e-01 FALSE  
## 40 FALSE 1.08218631 1.395849e-01 FALSE  
## 41 FALSE 1.08218631 1.395849e-01 FALSE  
## 42 FALSE 1.08218631 1.395849e-01 FALSE  
## 43 FALSE 1.08218631 1.395849e-01 FALSE  
## 44 FALSE 1.08218631 1.395849e-01 FALSE  
## 45 FALSE 1.08218631 1.395849e-01 FALSE  
## 46 FALSE 1.08218631 1.395849e-01 FALSE  
## 47 FALSE 1.08218631 1.395849e-01 FALSE  
## 48 FALSE 1.08218631 1.395849e-01 FALSE  
## 49 FALSE 1.08218631 1.395849e-01 FALSE  
## 50 FALSE 1.08218631 1.395849e-01 FALSE  
## 51 FALSE 1.08218631 1.395849e-01 FALSE  
## 52 FALSE 1.08218631 1.395849e-01 FALSE  
## 53 FALSE 1.08218631 1.395849e-01 FALSE  
## 54 FALSE 1.08218631 1.395849e-01 FALSE  
## 55 FALSE 1.08218631 1.395849e-01 FALSE  
## 56 FALSE 1.08218631 1.395849e-01 FALSE  
## 57 FALSE 1.08218631 1.395849e-01 FALSE  
## 58 FALSE 1.08218631 1.395849e-01 FALSE  
## 59 FALSE 1.08218631 1.395849e-01 FALSE  
## 60 FALSE 1.08218631 1.395849e-01 FALSE  
## 61 FALSE 1.08218631 1.395849e-01 FALSE  
## 62 FALSE 1.08218631 1.395849e-01 FALSE  
## 63 FALSE 1.08218631 1.395849e-01 FALSE  
## 64 FALSE 1.08218631 1.395849e-01 FALSE  
## 65 FALSE 1.08218631 1.395849e-01 FALSE  
## 66 FALSE 1.08218631 1.395849e-01 FALSE  
## 67 FALSE 1.08218631 1.395849e-01 FALSE  
## 68 FALSE 1.08218631 1.395849e-01 FALSE  
## 69 FALSE 1.08218631 1.395849e-01 FALSE  
## 70 FALSE 1.08218631 1.395849e-01 FALSE  
## 71 FALSE 1.08218631 1.395849e-01 FALSE  
## 72 FALSE 1.08218631 1.395849e-01 FALSE  
## 73 FALSE 1.08218631 1.395849e-01 FALSE  
## 74 FALSE 1.08218631 1.395849e-01 FALSE  
## 75 FALSE 1.08218631 1.395849e-01 FALSE  
## 76 FALSE 1.08218631 1.395849e-01 FALSE  
## 77 FALSE 1.08218631 1.395849e-01 FALSE  
## 78 FALSE 1.08218631 1.395849e-01 FALSE  
## 79 FALSE 1.08218631 1.395849e-01 FALSE  
## 80 FALSE 1.08218631 1.395849e-01 FALSE  
## 81 FALSE 1.08218631 1.395849e-01 FALSE  
## 82 FALSE 1.08218631 1.395849e-01 FALSE  
## 83 FALSE 1.08218631 1.395849e-01 FALSE  
## 84 FALSE 1.08218631 1.395849e-01 FALSE  
## 85 FALSE 1.08218631 1.395849e-01 FALSE  
## 86 FALSE 1.08218631 1.395849e-01 FALSE  
## 87 FALSE 1.08218631 1.395849e-01 FALSE  
## 88 FALSE 1.02200235 1.533899e-01 FALSE  
## 89 FALSE 0.98519951 1.622630e-01 FALSE  
## 90 FALSE 0.97805131 1.640245e-01 FALSE  
## 91 FALSE 0.97805131 1.640245e-01 FALSE  
## 92 FALSE 0.97805131 1.640245e-01 FALSE  
## 93 FALSE 0.97805131 1.640245e-01 FALSE  
## 94 FALSE 0.97805131 1.640245e-01 FALSE  
## 95 FALSE 0.97805131 1.640245e-01 FALSE  
## 96 FALSE 0.97805131 1.640245e-01 FALSE  
## 97 FALSE 0.97805131 1.640245e-01 FALSE  
## 98 FALSE 0.97805131 1.640245e-01 FALSE  
## 99 FALSE 0.97805131 1.640245e-01 FALSE  
## 100 FALSE 0.97805131 1.640245e-01 FALSE  
## 101 FALSE 0.97805131 1.640245e-01 FALSE  
## 102 FALSE 0.97805131 1.640245e-01 FALSE  
## 103 FALSE 0.97805131 1.640245e-01 FALSE  
## 104 FALSE 0.97805131 1.640245e-01 FALSE  
## 105 FALSE 0.97805131 1.640245e-01 FALSE  
## 106 FALSE 0.97805131 1.640245e-01 FALSE  
## 107 FALSE 0.97805131 1.640245e-01 FALSE  
## 108 FALSE 0.97805131 1.640245e-01 FALSE  
## 109 FALSE 0.97805131 1.640245e-01 FALSE  
## 110 FALSE 0.97805131 1.640245e-01 FALSE  
## 111 FALSE 0.97805131 1.640245e-01 FALSE  
## 112 FALSE 0.97805131 1.640245e-01 FALSE  
## 113 FALSE 0.97805131 1.640245e-01 FALSE  
## 114 FALSE 0.97805131 1.640245e-01 FALSE  
## 115 FALSE 0.97805131 1.640245e-01 FALSE  
## 116 FALSE 0.97805131 1.640245e-01 FALSE  
## 117 FALSE 0.97805131 1.640245e-01 FALSE  
## 118 FALSE 0.97805131 1.640245e-01 FALSE  
## 119 FALSE 0.97805131 1.640245e-01 FALSE  
## 120 FALSE 0.97805131 1.640245e-01 FALSE  
## 121 FALSE 0.97805131 1.640245e-01 FALSE  
## 122 FALSE 0.97805131 1.640245e-01 FALSE  
## 123 FALSE 0.97805131 1.640245e-01 FALSE  
## 124 FALSE 0.97805131 1.640245e-01 FALSE  
## 125 FALSE 0.97805131 1.640245e-01 FALSE  
## 126 FALSE 0.97805131 1.640245e-01 FALSE  
## 127 FALSE 0.97805131 1.640245e-01 FALSE  
## 128 FALSE 0.97805131 1.640245e-01 FALSE  
## 129 FALSE 0.97805131 1.640245e-01 FALSE  
## 130 FALSE 0.97805131 1.640245e-01 FALSE  
## 131 FALSE 0.97805131 1.640245e-01 FALSE  
## 132 FALSE 0.97805131 1.640245e-01 FALSE  
## 133 FALSE 0.97805131 1.640245e-01 FALSE  
## 134 FALSE 0.97805131 1.640245e-01 FALSE  
## 135 FALSE 0.96958356 1.661271e-01 FALSE  
## 136 FALSE 0.94255302 1.729548e-01 FALSE  
## 137 FALSE 0.92598829 1.772260e-01 FALSE  
## 138 FALSE 0.86608685 1.932213e-01 FALSE  
## 139 FALSE 0.84235673 1.997942e-01 FALSE  
## 140 FALSE 0.82334032 2.051572e-01 FALSE  
## 141 FALSE 0.81717165 2.069152e-01 FALSE  
## 142 FALSE 0.81335851 2.080063e-01 FALSE  
## 143 FALSE 0.78873352 2.151339e-01 FALSE  
## 144 FALSE 0.78418792 2.164649e-01 FALSE  
## 145 FALSE 0.78151112 2.172510e-01 FALSE  
## 146 FALSE 0.76852374 2.210880e-01 FALSE  
## 147 FALSE 0.76852374 2.210880e-01 FALSE  
## 148 FALSE 0.76852374 2.210880e-01 FALSE  
## 149 FALSE 0.76852374 2.210880e-01 FALSE  
## 150 FALSE 0.73302415 2.317718e-01 FALSE  
## 151 FALSE 0.72190706 2.351758e-01 FALSE  
## 152 FALSE 0.72190706 2.351758e-01 FALSE  
## 153 FALSE 0.72190706 2.351758e-01 FALSE  
## 154 FALSE 0.72190706 2.351758e-01 FALSE  
## 155 FALSE 0.68778139 2.457952e-01 FALSE  
## 156 FALSE 0.68773694 2.458092e-01 FALSE  
## 157 FALSE 0.66050863 2.544637e-01 FALSE  
## 158 FALSE 0.62290766 2.666726e-01 FALSE  
## 159 FALSE 0.60103717 2.739076e-01 FALSE  
## 160 FALSE 0.59851171 2.747493e-01 FALSE  
## 161 FALSE 0.58153792 2.804390e-01 FALSE  
## 162 FALSE 0.57238263 2.835314e-01 FALSE  
## 163 FALSE 0.56544118 2.858869e-01 FALSE  
## 164 FALSE 0.55345353 2.899764e-01 FALSE  
## 165 FALSE 0.55313615 2.900851e-01 FALSE  
## 166 FALSE 0.54776142 2.919279e-01 FALSE  
## 167 FALSE 0.52565152 2.995652e-01 FALSE  
## 168 FALSE 0.49883888 3.089464e-01 FALSE  
## 169 FALSE 0.47670885 3.167847e-01 FALSE  
## 170 FALSE 0.46705013 3.202320e-01 FALSE  
## 171 FALSE 0.46705013 3.202320e-01 FALSE  
## 172 FALSE 0.44115463 3.295505e-01 FALSE  
## 173 FALSE 0.42608938 3.350214e-01 FALSE  
## 174 FALSE 0.41747992 3.381637e-01 FALSE  
## 175 FALSE 0.41120941 3.404595e-01 FALSE  
## 176 FALSE 0.40473564 3.428359e-01 FALSE  
## 177 FALSE 0.40473564 3.428359e-01 FALSE  
## 178 FALSE 0.38941587 3.484843e-01 FALSE  
## 179 FALSE 0.37668984 3.532021e-01 FALSE  
## 180 FALSE 0.37668984 3.532021e-01 FALSE  
## 181 FALSE 0.37668984 3.532021e-01 FALSE  
## 182 FALSE 0.37668984 3.532021e-01 FALSE  
## 183 FALSE 0.34059536 3.667041e-01 FALSE  
## 184 FALSE 0.33708406 3.680268e-01 FALSE  
## 185 FALSE 0.29271897 3.848685e-01 FALSE  
## 186 FALSE 0.29271897 3.848685e-01 FALSE  
## 187 FALSE 0.29271897 3.848685e-01 FALSE  
## 188 FALSE 0.29271897 3.848685e-01 FALSE  
## 189 FALSE 0.29075559 3.856191e-01 FALSE  
## 190 FALSE 0.29075559 3.856191e-01 FALSE  
## 191 FALSE 0.29075559 3.856191e-01 FALSE  
## 192 FALSE 0.29075559 3.856191e-01 FALSE  
## 193 FALSE 0.29075559 3.856191e-01 FALSE  
## 194 FALSE 0.29075559 3.856191e-01 FALSE  
## 195 FALSE 0.29075559 3.856191e-01 FALSE  
## 196 FALSE 0.26958889 3.937383e-01 FALSE  
## 197 FALSE 0.25038492 4.011448e-01 FALSE  
## 198 FALSE 0.23872928 4.056578e-01 FALSE  
## 199 FALSE 0.23872928 4.056578e-01 FALSE  
## 200 FALSE 0.23872928 4.056578e-01 FALSE  
## 201 FALSE 0.23872928 4.056578e-01 FALSE  
## 202 FALSE 0.23872928 4.056578e-01 FALSE  
## 203 FALSE 0.23872928 4.056578e-01 FALSE  
## 204 FALSE 0.23872928 4.056578e-01 FALSE  
## 205 FALSE 0.23333057 4.077524e-01 FALSE  
## 206 FALSE 0.23203602 4.082550e-01 FALSE  
## 207 FALSE 0.20290394 4.196051e-01 FALSE  
## 208 FALSE 0.18336969 4.272540e-01 FALSE  
## 209 FALSE 0.17728659 4.296416e-01 FALSE  
## 210 FALSE 0.14430723 4.426289e-01 FALSE  
## 211 FALSE 0.14430723 4.426289e-01 FALSE  
## 212 FALSE 0.14112007 4.438875e-01 FALSE  
## 213 FALSE 0.12651466 4.496623e-01 FALSE  
## 214 FALSE 0.12196187 4.514646e-01 FALSE  
## 215 FALSE 0.10908601 4.565671e-01 FALSE  
## 216 FALSE 0.09447126 4.623674e-01 FALSE  
## 217 FALSE 0.09447126 4.623674e-01 FALSE  
## 218 FALSE 0.09447126 4.623674e-01 FALSE  
## 219 FALSE 0.09447126 4.623674e-01 FALSE  
## 220 FALSE 0.09117960 4.636749e-01 FALSE  
## 221 FALSE 0.07713546 4.692579e-01 FALSE  
## 222 FALSE 0.07713546 4.692579e-01 FALSE  
## 223 FALSE 0.03876118 4.845404e-01 FALSE  
## 224 FALSE 0.03876118 4.845404e-01 FALSE  
## 225 FALSE 0.03824897 4.847446e-01 FALSE  
## 226 FALSE 0.03275917 4.869333e-01 FALSE  
## 227 FALSE 0.02930070 4.883124e-01 FALSE  
## 228 FALSE 0.02537515 4.898779e-01 FALSE  
## 229 FALSE 0.02537515 4.898779e-01 FALSE  
## 230 FALSE 0.02537515 4.898779e-01 FALSE  
## 231 FALSE 0.02537515 4.898779e-01 FALSE  
## 232 FALSE 0.02071872 4.917350e-01 FALSE  
## 233 FALSE 0.02071872 4.917350e-01 FALSE  
##   
##   
## Summary statistics:  
##   
## Total Chi squared = 244.1559   
## Total degrees of freedom = 520   
## p = 0.0002973899   
## Sum of counts = 2140   
##   
## Levels:  
##   
## Status Variety X   
## 2 2 174

r8 <- cfa::cfa(configs, counts)  
# save to disc  
config\_words <- r8$table %>%  
 as.data.frame() %>%  
 dplyr::mutate(expected = round(expected, 1),  
 Q = round(Q, 3),  
 chisq = round(chisq, 3),  
 z = round(z, 3),  
 p.z = round(p.z, 3),  
 label = stringr::str\_remove\_all(label, "Xmuch ")) %>%  
 dplyr::select(-p.chisq, -sig.chisq)  
# save to disc  
write.table(config\_words,   
 here::here("tables", "xmuch\_wordconfigs.txt"),   
 sep = "\t",   
 row.names = F)  
# save to disc  
words <- r8$table %>%  
 as.data.frame() %>%  
 dplyr::filter(stringr::str\_detect(label, "control", negate = TRUE),  
 sig.z == T) %>%  
 dplyr::pull(label) %>%  
 stringr::str\_remove\_all("Xmuch ")  
words

## [1] "phrase Australia other"

The words that are attracted to the X-much structure in Ireland are less negative compared to the words that are attracted to the x-much construction in Oz. Compare *excuses*, *coincidence*, or *agenda* (Ireland) to *jealous*, *paranoid*, *hypocrite*, *creepy*, *threatened*, or *entitled* (Australia).

# Outro

sessionInfo()

## R version 4.1.1 (2021-08-10)  
## Platform: x86\_64-w64-mingw32/x64 (64-bit)  
## Running under: Windows 10 x64 (build 19043)  
##   
## 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] grid stats graphics grDevices utils datasets methods   
## [8] base   
##   
## other attached packages:  
## [1] effectsize\_0.5 vcd\_1.4-9 cfa\_0.10-0 tidytext\_0.3.2   
## [5] quanteda\_3.2.0 flextable\_0.6.10 readxl\_1.3.1 here\_1.0.1   
## [9] forcats\_0.5.1 stringr\_1.4.0 dplyr\_1.0.7 purrr\_0.3.4   
## [13] readr\_2.1.1 tidyr\_1.1.4 tibble\_3.1.6 ggplot2\_3.3.5   
## [17] tidyverse\_1.3.1   
##   
## loaded via a namespace (and not attached):  
## [1] fs\_1.5.2 lubridate\_1.8.0 RColorBrewer\_1.1-2 insight\_0.14.5   
## [5] httr\_1.4.2 rprojroot\_2.0.2 SnowballC\_0.7.0 tools\_4.1.1   
## [9] backports\_1.4.1 utf8\_1.2.2 R6\_2.5.1 DBI\_1.1.2   
## [13] colorspace\_2.0-2 withr\_2.4.3 tidyselect\_1.1.1 emmeans\_1.7.1-1   
## [17] compiler\_4.1.1 performance\_0.8.0 cli\_3.1.0 rvest\_1.0.2   
## [21] xml2\_1.3.3 sandwich\_3.0-1 officer\_0.4.1 bayestestR\_0.11.5   
## [25] scales\_1.1.1 mvtnorm\_1.1-3 lmtest\_0.9-39 systemfonts\_1.0.3   
## [29] digest\_0.6.28 rmarkdown\_2.11 base64enc\_0.1-3 pkgconfig\_2.0.3   
## [33] htmltools\_0.5.2 highr\_0.9 dbplyr\_2.1.1 fastmap\_1.1.0   
## [37] rlang\_0.4.12 rstudioapi\_0.13 generics\_0.1.1 zoo\_1.8-9   
## [41] jsonlite\_1.7.2 zip\_2.2.0 tokenizers\_0.2.1 magrittr\_2.0.1   
## [45] parameters\_0.15.0 Matrix\_1.4-0 Rcpp\_1.0.7 munsell\_0.5.0   
## [49] fansi\_0.5.0 gdtools\_0.2.3 lifecycle\_1.0.1 multcomp\_1.4-17   
## [53] stringi\_1.7.6 yaml\_2.2.1 MASS\_7.3-54 crayon\_1.4.2   
## [57] lattice\_0.20-45 splines\_4.1.1 haven\_2.4.3 hms\_1.1.1   
## [61] knitr\_1.37 pillar\_1.6.4 uuid\_1.0-3 estimability\_1.3   
## [65] codetools\_0.2-18 stopwords\_2.3 fastmatch\_1.1-3 reprex\_2.0.1.9000   
## [69] glue\_1.6.0 evaluate\_0.14 data.table\_1.14.2 RcppParallel\_5.1.4  
## [73] modelr\_0.1.8 vctrs\_0.3.8 tzdb\_0.2.0 cellranger\_1.1.0   
## [77] gtable\_0.3.0 datawizard\_0.2.1 assertthat\_0.2.1 xfun\_0.29   
## [81] xtable\_1.8-4 broom\_0.7.10 coda\_0.19-4 janeaustenr\_0.1.5   
## [85] survival\_3.2-13 TH.data\_1.1-0 ellipsis\_0.3.2

# References

Ben-Shachar, Mattan S., Daniel Lüdecke, and Dominique Makowski. 2020. “effectsize: Estimation of Effect Size Indices and Standardized Parameters.” *Journal of Open Source Software* 5 (56): 2815. <https://doi.org/10.21105/joss.02815>.

Cochran, William G. 1952. “The χ2 Test of Goodness of Fit.” *The Annals of Mathematical Statistics* 23 (3): 315–45.

Lienert, Gustav A., and Joachim Krauth. 1975. “Configural Frequency Analysis as a Statistical Tool for Defining Types.” *Educational and Psychological Measurement* 35 (2): 231–38.

Mair, Patrick, and Stefan Funke. 2017. *Cfa: Configural Frequency Analysis (Cfa)*. <https://CRAN.R-project.org/package=cfa>.

R Core Team. 2021. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. <https://www.R-project.org/>.

Wickham, Hadley, Mara Averick, Jennifer Bryan, Winston Chang, Lucy D’Agostino McGowan, Romain François, Garrett Grolemund, et al. 2019. “Welcome to the tidyverse.” *Journal of Open Source Software* 4 (43): 1686. <https://doi.org/10.21105/joss.01686>.