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

Anonymous

2022-01-01

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("party")  
install.packages("partykit")  
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(party)

## Lade nötiges Paket: mvtnorm

## Lade nötiges Paket: modeltools

## Lade nötiges Paket: stats4

## Lade nötiges Paket: strucchange

## Lade nötiges Paket: zoo

##   
## Attache Paket: 'zoo'

## Das folgende Objekt ist maskiert 'package:quanteda':  
##   
## index

## Die folgenden Objekte sind maskiert von 'package:base':  
##   
## as.Date, as.Date.numeric

## Lade nötiges Paket: sandwich

##   
## Attache Paket: 'strucchange'

## Das folgende Objekt ist maskiert 'package:stringr':  
##   
## boundary

library(partykit)

## Lade nötiges Paket: libcoin

##   
## Attache Paket: 'partykit'

## Die folgenden Objekte sind maskiert von 'package:party':  
##   
## cforest, ctree, ctree\_control, edge\_simple, mob, mob\_control,  
## node\_barplot, node\_bivplot, node\_boxplot, node\_inner, node\_surv,  
## node\_terminal, varimp

## Das folgende Objekt ist maskiert 'package:flextable':  
##   
## width

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 element emo control polarity  
## 1 1 cute Adjective word emotional test positive  
## 2 5 creative Adjective word emotional test positive  
## 3 7 hype Noun word emotional test negative  
## 4 8 twins Noun word nonemotional test nonemotional  
## 5 11 pregnancy craving Noun phrase nonemotional test nonemotional  
## 6 17 feasible Adjective word nonemotional test nonemotional  
## 7 26 airy fairy Noun phrase nonemotional test nonemotional  
## 8 27 ironic Adjective word nonemotional test nonemotional  
## 9 29 jealous Adjective word emotional test negative  
## 10 32 obsessed Adjective word nonemotional test nonemotional

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" ...  
## $ element : chr "word" "word" "word" "word" ...  
## $ emo : chr "emotional" "emotional" "emotional" "nonemotional" ...  
## $ control : chr "test" "test" "test" "test" ...  
## $ polarity: chr "positive" "positive" "negative" "nonemotional" ...

# Statistical Analysis

## Single vs Multiword

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

phrasetb <- xmuch %>%  
 dplyr::filter(control == "test") %>%  
 dplyr::select(variety, element) %>%  
 dplyr::group\_by(variety, element) %>%  
 dplyr::summarise(freq = n()) %>%  
 dplyr::group\_by(variety) %>%  
 dplyr::mutate(total = sum(freq)) %>%  
 dplyr::ungroup() %>%  
 dplyr::rowwise() %>%  
 dplyr::mutate(percent = round(freq/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 element freq 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(element, freq) %>%  
 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")

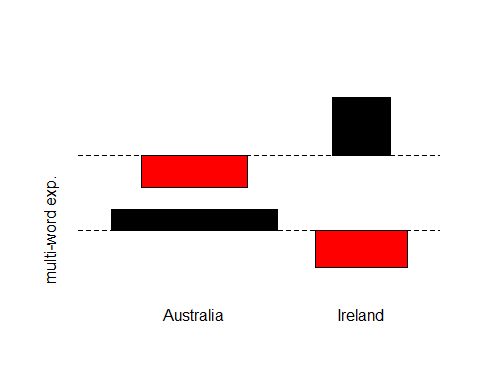
## Phi | 95% CI  
## -------------------  
## 0.02 | [0.00, 1.00]  
##   
## - One-sided CIs: upper bound fixed at (1).

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, control, polarity) %>%  
 dplyr::summarise(Freq = n()) %>%  
 dplyr::group\_by(variety, control) %>%  
 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', 'control'. You can override using the `.groups` argument.

# inspect  
emo\_tab

## # A tibble: 4 x 5  
## variety control negative nonemotional positive  
## <chr> <chr> <chr> <chr> <chr>   
## 1 Australia control 19 (1.9) 930 (93) 51 (5.1)  
## 2 Australia test 20 (18.3) 80 (73.4) 9 (8.3)   
## 3 Ireland control 29 (2.9) 919 (91.9) 52 (5.2)  
## 4 Ireland test 6 (18.2) 26 (78.8) 1 (3)

Calculate totals

# au  
austot <- 19+20+930+80+51+9  
negau <- round(39/austot\*100, 1)  
nonau <- round(1010/austot\*100, 1)  
posau <- round(60/austot\*100, 1)  
# ire  
iretot <- 29+6+919+26+52+1  
negire <- round(35/iretot\*100, 1)  
nonire <- round(945/iretot\*100, 1)  
posire <- round(53/iretot\*100, 1)  
# results  
negau

## [1] 3.5

nonau

## [1] 91.1

posau

## [1] 5.4

negire

## [1] 3.4

nonire

## [1] 91.5

posire

## [1] 5.1

xmdtb <- xmuch %>%  
 dplyr::select(variety, x, pos, emo, polarity, control) %>%  
 dplyr::rename(word = x)  
# inspect  
head(xmdtb)

## variety word pos emo polarity control  
## 1 Australia cute Adjective emotional positive test  
## 2 Australia creative Adjective emotional positive test  
## 3 Australia hype Noun emotional negative test  
## 4 Australia twins Noun nonemotional nonemotional test  
## 5 Australia pregnancy craving Noun nonemotional nonemotional test  
## 6 Australia feasible Adjective nonemotional nonemotional test

* 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, control) %>%  
 dplyr::summarise(freq = n())

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

## # A tibble: 4 x 3  
## # Groups: variety [2]  
## variety control freq  
## <chr> <chr> <int>  
## 1 Australia control 1000  
## 2 Australia test 109  
## 3 Ireland control 1000  
## 4 Ireland test 33

### 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(emo, control) %>%  
 dplyr::summarise(freq = n())

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

# inspect  
head(emo\_tb1)

## # A tibble: 4 x 3  
## # Groups: emo [2]  
## emo control freq  
## <chr> <chr> <int>  
## 1 emotional control 151  
## 2 emotional test 36  
## 3 nonemotional control 1849  
## 4 nonemotional test 106

Perform x2-test

emo\_x2 <- emo\_tb1 %>%  
 tidyr::spread(control, freq) %>%  
 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 test  
## emotional 151 36  
## nonemotional 1849 106

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

##   
## Pearson's Chi-squared test with Yates' continuity correction  
##   
## data: emo\_x2  
## X-squared = 50.524, df = 1, p-value = 1.177e-12

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

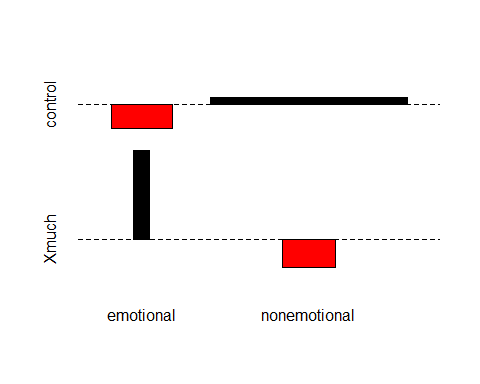
## Phi | 95% CI  
## -------------------  
## 0.16 | [0.12, 1.00]  
##   
## - One-sided CIs: upper bound fixed at (1).

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", "Xmuch")  
  
# 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(control == "test") %>%  
 dplyr::group\_by(emo, variety) %>%  
 dplyr::summarise(freq = n())

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

# inspect  
head(emo\_tb2)

## # A tibble: 4 x 3  
## # Groups: emo [2]  
## emo variety freq  
## <chr> <chr> <int>  
## 1 emotional Australia 29  
## 2 emotional Ireland 7  
## 3 nonemotional Australia 80  
## 4 nonemotional Ireland 26

Perform x2-test

emo\_x22 <- emo\_tb2 %>%  
 tidyr::spread(variety, freq) %>%  
 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 29 7  
## nonemotional 80 26

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

##   
## Pearson's Chi-squared test with Yates' continuity correction  
##   
## data: emo\_x22  
## X-squared = 0.15651, df = 1, p-value = 0.6924

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

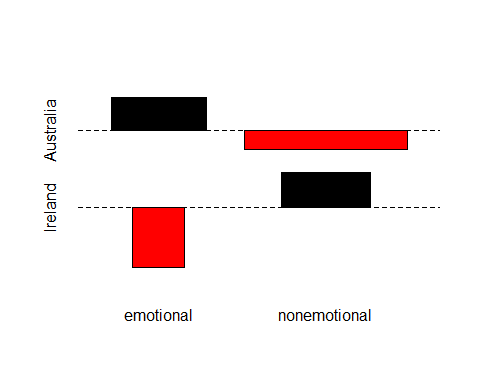
## Phi | 95% CI  
## -------------------  
## 0.05 | [0.00, 1.00]  
##   
## - One-sided CIs: upper bound fixed at (1).

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(emo, variety, control) %>%  
 dplyr::summarise(freq = n())

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

# inspect  
emo\_tb3

## # A tibble: 8 x 4  
## # Groups: emo, variety [4]  
## emo variety control freq  
## <chr> <chr> <chr> <int>  
## 1 emotional Australia control 70  
## 2 emotional Australia test 29  
## 3 emotional Ireland control 81  
## 4 emotional Ireland test 7  
## 5 nonemotional Australia control 930  
## 6 nonemotional Australia test 80  
## 7 nonemotional Ireland control 919  
## 8 nonemotional Ireland test 26

configs <- emo\_tb3 %>%  
 dplyr::select(variety, emo, control)  
counts = emo\_tb3$freq  
cfa::cfa(configs, counts)

##   
## \*\*\* Analysis of configuration frequencies (CFA) \*\*\*  
##   
## label n expected Q chisq  
## 1 Australia emotional test 29 6.418338 0.0105740102 79.4491495  
## 2 Ireland nonemotional test 26 62.502371 0.0175534566 21.3179610  
## 3 Australia emotional control 70 90.399123 0.0099430269 4.6031885  
## 4 Australia nonemotional test 80 67.100803 0.0062167824 2.4796913  
## 5 Ireland nonemotional control 919 880.315089 0.0306613091 1.6999849  
## 6 Australia nonemotional control 930 945.081736 0.0126004731 0.2406763  
## 7 Ireland emotional test 7 5.978488 0.0004782313 0.1745404  
## 8 Ireland emotional control 81 84.204052 0.0015570310 0.1219175  
## p.chisq sig.chisq z p.z sig.z  
## 1 0.000000e+00 TRUE 8.7291534 0.00000000 TRUE  
## 2 3.890690e-06 TRUE -4.7501993 0.99999898 TRUE  
## 3 3.191255e-02 FALSE -2.2459974 0.98764791 FALSE  
## 4 1.153249e-01 FALSE 1.5379454 0.06203098 FALSE  
## 5 1.922900e-01 FALSE 1.6768993 0.04678109 FALSE  
## 6 6.237181e-01 FALSE -0.6780452 0.75112849 FALSE  
## 7 6.761078e-01 FALSE 0.2135875 0.41543438 FALSE  
## 8 7.269640e-01 FALSE -0.4118311 0.65976839 FALSE  
##   
##   
## Summary statistics:  
##   
## Total Chi squared = 110.0871   
## Total degrees of freedom = 4   
## p = 0   
## Sum of counts = 2142   
##   
## Levels:  
##   
## variety emo control   
## 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::rename(Freq = freq,  
 Emotionality = emo,  
 Variety = variety,  
 Data = control) %>%  
 dplyr::mutate\_if(is.character, factor) %>%  
 dplyr::mutate(Data = ifelse(Data == "test", "X-much", "control"))  
# inspect  
head(emo\_vis)

## # A tibble: 6 x 4  
## Emotionality Variety Data Freq  
## <fct> <fct> <chr> <int>  
## 1 emotional Australia control 70  
## 2 emotional Australia X-much 29  
## 3 emotional Ireland control 81  
## 4 emotional Ireland X-much 7  
## 5 nonemotional Australia control 930  
## 6 nonemotional Australia X-much 80

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

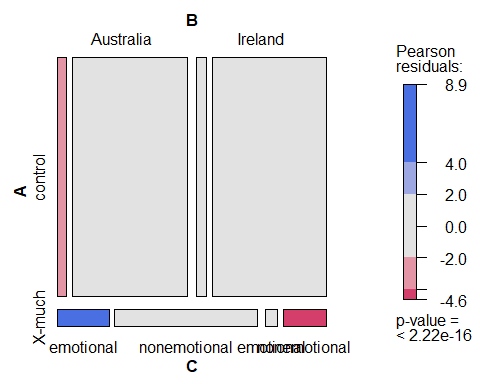
## , , emotional  
##   
## Australia Ireland  
## control 70 81  
## X-much 29 7  
##   
## , , nonemotional  
##   
## Australia Ireland  
## control 930 919  
## X-much 80 26

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::mutate(polarity = ifelse(polarity == "nonemotional", "neutral", polarity)) %>%  
 dplyr::group\_by(polarity, control) %>%  
 dplyr::summarise(freq = n())

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

# inspect  
pol\_tb1

## # A tibble: 6 x 3  
## # Groups: polarity [3]  
## polarity control freq  
## <chr> <chr> <int>  
## 1 negative control 48  
## 2 negative test 26  
## 3 neutral control 1849  
## 4 neutral test 106  
## 5 positive control 103  
## 6 positive test 10

Perform x2-test

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

## polarity control test   
## [1,] "negative" " 48" " 26"  
## [2,] "neutral" "1849" "106"  
## [3,] "positive" " 103" " 10"

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

## control test  
## negative 48 26  
## neutral 1849 106  
## positive 103 10

# 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 = 102.65, 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

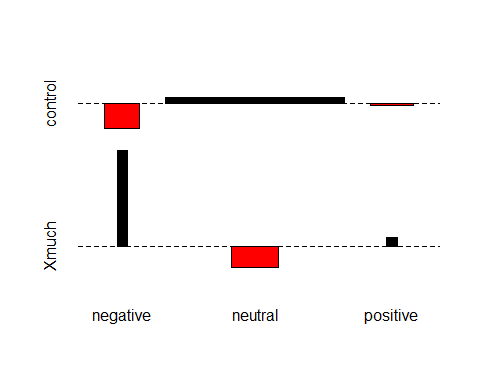
## Phi | 95% CI  
## -------------------  
## 0.22 | [0.18, 1.00]  
##   
## - One-sided CIs: upper bound fixed at (1).

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", "Xmuch")  
  
# 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(control == "test") %>%  
 dplyr::mutate(polarity = ifelse(polarity == "nonemotional", "neutral", polarity)) %>%  
 dplyr::group\_by(polarity, variety) %>%  
 dplyr::summarise(freq = 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 freq  
## <chr> <chr> <int>  
## 1 negative Australia 20  
## 2 negative Ireland 6  
## 3 neutral Australia 80  
## 4 neutral Ireland 26  
## 5 positive Australia 9  
## 6 positive Ireland 1

Perform x2-test

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

## polarity Australia Ireland  
## [1,] "negative" "20" " 6"   
## [2,] "neutral" "80" "26"   
## [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 20 6  
## neutral 80 26  
## 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 = 1.0817, df = 2, p-value = 0.5823

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

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

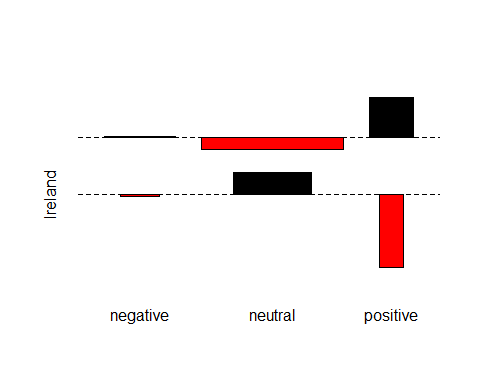
## Phi | 95% CI  
## -------------------  
## 0.09 | [0.00, 1.00]  
##   
## - One-sided CIs: upper bound fixed at (1).

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::mutate(polarity = ifelse(polarity == "nonemotional", "neutral", polarity)) %>%  
 dplyr::group\_by(polarity, variety, control) %>%  
 dplyr::summarise(freq = n())

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

# inspect  
pol\_tb1

## # A tibble: 12 x 4  
## # Groups: polarity, variety [6]  
## polarity variety control freq  
## <chr> <chr> <chr> <int>  
## 1 negative Australia control 19  
## 2 negative Australia test 20  
## 3 negative Ireland control 29  
## 4 negative Ireland test 6  
## 5 neutral Australia control 930  
## 6 neutral Australia test 80  
## 7 neutral Ireland control 919  
## 8 neutral Ireland test 26  
## 9 positive Australia control 51  
## 10 positive Australia test 9  
## 11 positive Ireland control 52  
## 12 positive Ireland test 1

configs <- pol\_tb1 %>%  
 dplyr::select(variety, polarity, control)  
counts = pol\_tb1$freq  
cfa::cfa(configs, counts)

##   
## \*\*\* Analysis of configuration frequencies (CFA) \*\*\*  
##   
## label n expected Q chisq  
## 1 Australia negative test 20 2.539877 0.0081609949 120.02782121  
## 2 Ireland neutral test 26 62.502371 0.0175534566 21.31796096  
## 3 Australia negative control 19 35.772915 0.0079634883 7.86434858  
## 4 Australia positive test 9 3.878461 0.0023953452 6.76303460  
## 5 Ireland negative test 6 2.365819 0.0016985059 5.58253862  
## 6 Australia neutral test 80 67.100803 0.0062167824 2.47969127  
## 7 Ireland positive test 1 3.612669 0.0012217941 1.88947269  
## 8 Ireland neutral control 919 880.315089 0.0306613091 1.69998488  
## 9 Ireland negative control 29 33.321390 0.0020493353 0.56043304  
## 10 Australia positive control 51 54.626208 0.0017372106 0.24071565  
## 11 Australia neutral control 930 945.081736 0.0126004731 0.24067630  
## 12 Ireland positive control 52 50.882662 0.0005343256 0.02453573  
## p.chisq sig.chisq z p.z sig.z  
## 1 0.000000e+00 TRUE 10.9557209 0.000000e+00 TRUE  
## 2 2.348895e-05 TRUE 4.6171377 1.945345e-06 TRUE  
## 3 1.960101e-02 FALSE 2.8043446 2.520949e-03 TRUE  
## 4 3.399583e-02 FALSE 2.6005835 4.653268e-03 FALSE  
## 5 6.134330e-02 FALSE 2.3627396 9.070204e-03 FALSE  
## 6 2.894289e-01 FALSE 1.5747035 5.766244e-02 FALSE  
## 7 3.887821e-01 FALSE 1.3745809 8.463070e-02 FALSE  
## 8 4.274182e-01 FALSE 1.3038347 9.614498e-02 FALSE  
## 9 7.556201e-01 FALSE 0.7486208 2.270429e-01 FALSE  
## 10 8.866031e-01 FALSE 0.4906278 3.118449e-01 FALSE  
## 11 8.866206e-01 FALSE 0.4905877 3.118590e-01 FALSE  
## 12 9.878071e-01 FALSE 0.1566389 4.377647e-01 FALSE  
##   
##   
## Summary statistics:  
##   
## Total Chi squared = 168.6912   
## Total degrees of freedom = 7   
## p = 0   
## Sum of counts = 2142   
##   
## Levels:  
##   
## variety polarity control   
## 2 3 2

### Visualization

* prepare data

pol\_vis <- pol\_tb1 %>%  
 dplyr::ungroup() %>%  
 dplyr::rename(Freq = freq,  
 Polarity = polarity,  
 Variety = variety) %>%  
 dplyr::mutate\_if(is.character, factor)  
# inspect  
pol\_vis

## # A tibble: 12 x 4  
## Polarity Variety control Freq  
## <fct> <fct> <fct> <int>  
## 1 negative Australia control 19  
## 2 negative Australia test 20  
## 3 negative Ireland control 29  
## 4 negative Ireland test 6  
## 5 neutral Australia control 930  
## 6 neutral Australia test 80  
## 7 neutral Ireland control 919  
## 8 neutral Ireland test 26  
## 9 positive Australia control 51  
## 10 positive Australia test 9  
## 11 positive Ireland control 52  
## 12 positive Ireland test 1

* convert to matrix

pol1 <- pol\_vis %>%  
 dplyr::filter(Polarity == "positive") %>%  
 dplyr::pull(Freq)  
pol2 <- pol\_vis %>%  
 dplyr::filter(Polarity == "neutral") %>%  
 dplyr::pull(Freq)  
pol3 <- pol\_vis %>%  
 dplyr::filter(Polarity == "negative") %>%  
 dplyr::pull(Freq)  
# 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

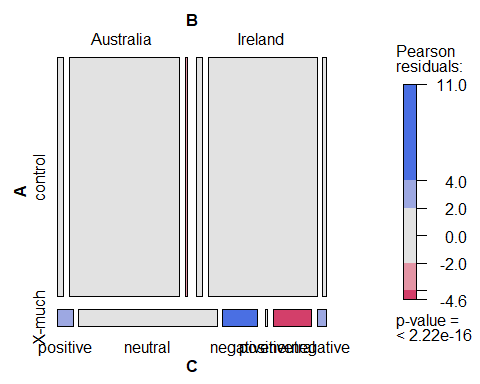
## , , positive  
##   
## Australia Ireland  
## control 51 52  
## X-much 9 1  
##   
## , , neutral  
##   
## Australia Ireland  
## control 930 919  
## X-much 80 26  
##   
## , , negative  
##   
## Australia Ireland  
## control 19 29  
## X-much 20 6

* 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()

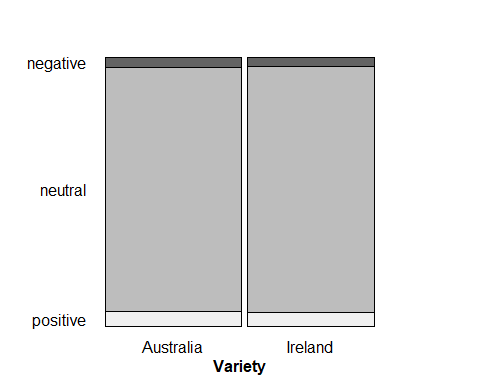
## png   
## 2

# 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, control, pos) %>%  
 dplyr::summarise(Freq = n()) %>%  
 dplyr::group\_by(variety, control) %>%  
 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', 'control'. You can override using the `.groups` argument.

# inspect  
pos\_tab

## # A tibble: 4 x 6  
## variety control Adjective Noun other Verb   
## <chr> <chr> <chr> <chr> <chr> <chr>   
## 1 Australia control 53 (5.3) 305 (30.5) 582 (58.2) 60 (6)   
## 2 Australia test 45 (41.3) 46 (42.2) 2 (1.8) 16 (14.7)  
## 3 Ireland control 76 (7.6) 319 (31.9) 542 (54.2) 63 (6.3)   
## 4 Ireland test 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, control) %>%  
 dplyr::summarise(freq = 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 control freq  
## <chr> <chr> <int>  
## 1 Adjective control 129  
## 2 Adjective test 53  
## 3 Noun control 624  
## 4 Noun test 69  
## 5 other control 1124  
## 6 other test 3  
## 7 Verb control 123  
## 8 Verb test 17

Perform x2-test

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

## pos control test  
## [1,] "Adjective" " 129" "53"  
## [2,] "Noun" " 624" "69"  
## [3,] "other" "1124" " 3"  
## [4,] "Verb" " 123" "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 test  
## Adjective 129 53  
## Noun 624 69  
## other 1124 3  
## Verb 123 17

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

##   
## Pearson's Chi-squared test  
##   
## data: pos\_x2  
## X-squared = 241.73, df = 3, p-value < 2.2e-16

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

## Phi | 95% CI  
## -------------------  
## 0.34 | [0.30, 1.00]  
##   
## - One-sided CIs: upper bound fixed at (1).

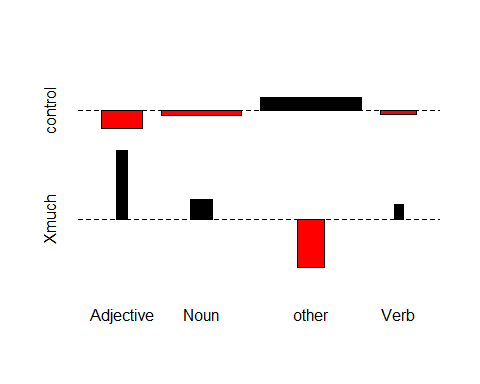
The word classes (parts-of-speech) of words in the x-much structure differ significantly from the word classes in the control data!

generate association plot

# adapt column names  
colnames(pos\_x2) <- c("control", "Xmuch")  
  
# 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(control == "test") %>%  
 dplyr::group\_by(pos, variety) %>%  
 dplyr::summarise(freq = 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 freq  
## <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, freq) %>%  
 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).

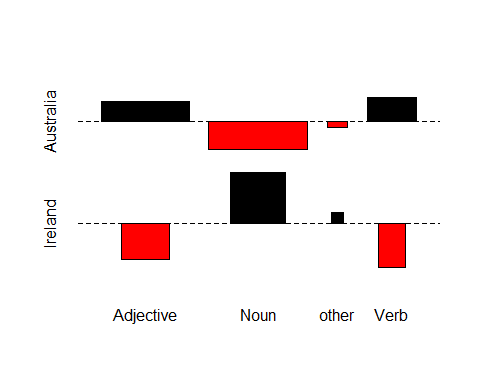
Australia and Ireland differ significantly with respect to the word classes (parts-of-speech) of words in the x-much structure!

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, control) %>%  
 dplyr::summarise(freq = 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 control freq  
## <chr> <chr> <chr> <int>  
## 1 Adjective Australia control 53  
## 2 Adjective Australia test 45  
## 3 Adjective Ireland control 76  
## 4 Adjective Ireland test 8  
## 5 Noun Australia control 305  
## 6 Noun Australia test 46  
## 7 Noun Ireland control 319  
## 8 Noun Ireland test 23  
## 9 other Australia control 582  
## 10 other Australia test 2  
## 11 other Ireland control 542  
## 12 other Ireland test 1  
## 13 Verb Australia control 60  
## 14 Verb Australia test 16  
## 15 Verb Ireland control 63  
## 16 Verb Ireland test 1

perform CFA

configs <- pos\_tb2 %>%  
 dplyr::select(variety, pos, control)  
counts = pos\_tb2$freq  
cfa::cfa(configs, counts)

##   
## \*\*\* Analysis of configuration frequencies (CFA) \*\*\*  
##   
## label n expected Q chisq  
## 1 Australia Adjective test 45 6.246724 0.0181450152 2.404166e+02  
## 2 Australia other test 2 38.681640 0.0174398894 3.478505e+01  
## 3 Ireland other test 1 36.030779 0.0166340412 3.405853e+01  
## 4 Australia Verb test 16 4.805173 0.0052380940 2.608109e+01  
## 5 Australia Noun test 46 23.785604 0.0104873216 2.074698e+01  
## 6 Australia Adjective control 53 87.982034 0.0170310262 1.390901e+01  
## 7 Ireland Verb test 1 4.475873 0.0016261211 2.699293e+00  
## 8 Australia Noun control 305 335.008513 0.0166068924 2.688024e+00  
## 9 Australia other control 582 544.811825 0.0232835279 2.538418e+00  
## 10 Ireland other control 542 507.475757 0.0211218910 2.348730e+00  
## 11 Australia Verb control 60 67.678488 0.0037016863 8.711656e-01  
## 12 Ireland Adjective test 8 5.818635 0.0010211516 8.177782e-01  
## 13 Ireland Adjective control 76 81.952607 0.0028895484 4.323661e-01  
## 14 Ireland Noun control 319 312.050310 0.0037977490 1.547769e-01  
## 15 Ireland Noun test 23 22.155572 0.0003983443 3.218417e-02  
## 16 Ireland Verb control 63 63.040467 0.0000194649 2.597627e-05  
## p.chisq sig.chisq z p.z sig.z  
## 1 0.000000e+00 TRUE 15.505373928 0.000000e+00 TRUE  
## 2 1.352526e-07 TRUE 5.897885025 1.840952e-09 TRUE  
## 3 1.925611e-07 TRUE 5.835968884 2.673944e-09 TRUE  
## 4 9.171711e-06 TRUE 5.106965276 1.636869e-07 TRUE  
## 5 1.188128e-04 TRUE 4.554884877 2.620714e-06 TRUE  
## 6 3.031663e-03 TRUE 3.729478058 9.593841e-05 TRUE  
## 7 4.403474e-01 FALSE 1.642952596 5.019637e-02 FALSE  
## 8 4.422664e-01 FALSE 1.639519388 5.055257e-02 FALSE  
## 9 4.683880e-01 FALSE 1.593241497 5.555301e-02 FALSE  
## 10 5.032499e-01 FALSE 1.532556563 6.269258e-02 FALSE  
## 11 8.323800e-01 FALSE 0.933362526 1.753164e-01 FALSE  
## 12 8.452102e-01 FALSE 0.904310897 1.829153e-01 FALSE  
## 13 9.334782e-01 FALSE 0.657545495 2.554151e-01 FALSE  
## 14 9.845367e-01 FALSE 0.393416983 3.470058e-01 FALSE  
## 15 9.984791e-01 FALSE 0.179399457 4.288120e-01 FALSE  
## 16 1.000000e+00 FALSE 0.005096692 4.979667e-01 FALSE  
##   
##   
## Summary statistics:  
##   
## Total Chi squared = 382.58   
## Total degrees of freedom = 10   
## p = 0   
## Sum of counts = 2142   
##   
## Levels:  
##   
## variety pos control   
## 2 4 2

### Visualization

* prepare data

pos\_vis <- pos\_tb2 %>%  
 dplyr::ungroup() %>%  
 dplyr::rename(Freq = freq,  
 POS = pos,  
 Variety = variety) %>%  
 dplyr::mutate\_if(is.character, factor)  
# inspect  
pos\_vis

## # A tibble: 16 x 4  
## POS Variety control Freq  
## <fct> <fct> <fct> <int>  
## 1 Adjective Australia control 53  
## 2 Adjective Australia test 45  
## 3 Adjective Ireland control 76  
## 4 Adjective Ireland test 8  
## 5 Noun Australia control 305  
## 6 Noun Australia test 46  
## 7 Noun Ireland control 319  
## 8 Noun Ireland test 23  
## 9 other Australia control 582  
## 10 other Australia test 2  
## 11 other Ireland control 542  
## 12 other Ireland test 1  
## 13 Verb Australia control 60  
## 14 Verb Australia test 16  
## 15 Verb Ireland control 63  
## 16 Verb Ireland test 1

* prepare data

pos1 <- pos\_vis %>%  
 dplyr::filter(POS == "Adjective") %>%  
 dplyr::pull(Freq)  
pos2 <- pos\_vis %>%  
 dplyr::filter(POS == "Noun") %>%  
 dplyr::pull(Freq)  
pos3 <- pos\_vis %>%  
 dplyr::filter(POS == "other") %>%  
 dplyr::pull(Freq)  
pos4 <- pos\_vis %>%  
 dplyr::filter(POS == "Verb") %>%  
 dplyr::pull(Freq)  
# 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

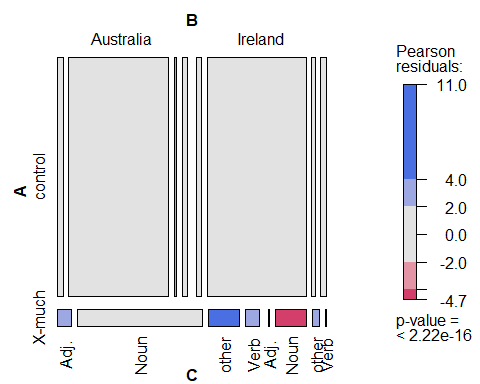
## , , Adj.  
##   
## Australia Ireland  
## control 51 52  
## X-much 9 1  
##   
## , , Noun  
##   
## Australia Ireland  
## control 930 919  
## X-much 80 26  
##   
## , , other  
##   
## Australia Ireland  
## control 19 29  
## X-much 20 6  
##   
## , , Verb  
##   
## Australia Ireland  
## control 51 52  
## X-much 9 1

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(control, x) %>%  
 dplyr::mutate(freq = n()) %>%  
 dplyr::mutate(x = ifelse(freq < 2, "other", x)) %>%  
 dplyr::ungroup() %>%  
 dplyr::group\_by(x, control) %>%  
 dplyr::summarise(freq = n())

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

configs <- wordcat\_tb %>%  
 dplyr::select(control, x)  
counts = wordcat\_tb$freq  
cfa::cfa(configs, counts)

##   
## \*\*\* Analysis of configuration frequencies (CFA) \*\*\*  
##   
## label n expected Q chisq p.chisq  
## 1 test jealous 6 0.3986898 2.621598e-03 78.694464400 1  
## 2 test paranoid 4 0.2657932 1.747623e-03 52.462976267 1  
## 3 test agenda 4 0.2657932 1.747623e-03 52.462976267 1  
## 4 test threatened 3 0.1993449 1.310677e-03 39.347232200 1  
## 5 test other 105 60.6672906 2.135145e-02 32.396190830 1  
## 6 test hypocrite 2 0.1328966 8.737574e-04 26.231488133 1  
## 7 test hipster 2 0.1328966 8.737574e-04 26.231488133 1  
## 8 test generalise 2 0.1328966 8.737574e-04 26.231488133 1  
## 9 test excuses 2 0.1328966 8.737574e-04 26.231488133 1  
## 10 test excited 2 0.1328966 8.737574e-04 26.231488133 1  
## 11 test entitled 2 0.1328966 8.737574e-04 26.231488133 1  
## 12 test creepy 2 0.1328966 8.737574e-04 26.231488133 1  
## 13 test coincidence 2 0.1328966 8.737574e-04 26.231488133 1  
## 14 test bitter 2 0.1328966 8.737574e-04 26.231488133 1  
## 15 test biased 2 0.1328966 8.737574e-04 26.231488133 1  
## 16 control other 808 852.3327094 3.450910e-02 2.305894285 1  
## 17 control the 142 132.5643425 4.707389e-03 0.671610707 1  
## 18 control to 63 58.8137576 2.014373e-03 0.297968130 1  
## 19 control and 58 54.1459991 1.850346e-03 0.274319866 1  
## 20 control of 48 44.8104820 1.524488e-03 0.227023337 1  
## 21 control a 43 40.1427234 1.362647e-03 0.203375073 1  
## 22 control in 34 31.7407581 1.073142e-03 0.160808197 1  
## 23 control is 31 28.9401029 9.771530e-04 0.146619239 1  
## 24 control that 28 26.1394478 8.814188e-04 0.132430280 1  
## 25 control i 25 23.3387927 7.859383e-04 0.118241322 1  
## 26 control it 24 22.4052410 7.541677e-04 0.113511669 1  
## 27 control with 18 16.8039307 5.641314e-04 0.085133752 1  
## 28 control this 17 15.8703790 5.325563e-04 0.080404099 1  
## 29 control on 16 14.9368273 5.010090e-04 0.075674446 1  
## 30 control for 16 14.9368273 5.010090e-04 0.075674446 1  
## 31 control you 14 13.0697239 4.379975e-04 0.066215140 1  
## 32 control was 14 13.0697239 4.379975e-04 0.066215140 1  
## 33 control have 14 13.0697239 4.379975e-04 0.066215140 1  
## 34 control by 14 13.0697239 4.379975e-04 0.066215140 1  
## 35 control are 14 13.0697239 4.379975e-04 0.066215140 1  
## 36 control if 12 11.2026205 3.750967e-04 0.056755834 1  
## 37 control be 12 11.2026205 3.750967e-04 0.056755834 1  
## 38 control an 11 10.2690688 3.436877e-04 0.052026182 1  
## 39 control they 10 9.3355171 3.123063e-04 0.047296529 1  
## 40 control he 10 9.3355171 3.123063e-04 0.047296529 1  
## 41 control would 9 8.4019654 2.809524e-04 0.042566876 1  
## 42 control we 9 8.4019654 2.809524e-04 0.042566876 1  
## 43 control not 9 8.4019654 2.809524e-04 0.042566876 1  
## 44 control from 9 8.4019654 2.809524e-04 0.042566876 1  
## 45 control will 8 7.4684137 2.496259e-04 0.037837223 1  
## 46 control were 8 7.4684137 2.496259e-04 0.037837223 1  
## 47 control when 7 6.5348620 2.183270e-04 0.033107570 1  
## 48 control one 7 6.5348620 2.183270e-04 0.033107570 1  
## 49 control but 7 6.5348620 2.183270e-04 0.033107570 1  
## 50 control as 7 6.5348620 2.183270e-04 0.033107570 1  
## 51 control all 7 6.5348620 2.183270e-04 0.033107570 1  
## 52 control your 6 5.6013102 1.870555e-04 0.028377917 1  
## 53 control which 6 5.6013102 1.870555e-04 0.028377917 1  
## 54 control what 6 5.6013102 1.870555e-04 0.028377917 1  
## 55 control there 6 5.6013102 1.870555e-04 0.028377917 1  
## 56 control their 6 5.6013102 1.870555e-04 0.028377917 1  
## 57 control or 6 5.6013102 1.870555e-04 0.028377917 1  
## 58 control do 6 5.6013102 1.870555e-04 0.028377917 1  
## 59 control than 5 4.6677585 1.558113e-04 0.023648264 1  
## 60 control so 5 4.6677585 1.558113e-04 0.023648264 1  
## 61 control radio 5 4.6677585 1.558113e-04 0.023648264 1  
## 62 control over 5 4.6677585 1.558113e-04 0.023648264 1  
## 63 control only 5 4.6677585 1.558113e-04 0.023648264 1  
## 64 control my 5 4.6677585 1.558113e-04 0.023648264 1  
## 65 control just 5 4.6677585 1.558113e-04 0.023648264 1  
## 66 control its 5 4.6677585 1.558113e-04 0.023648264 1  
## 67 control has 5 4.6677585 1.558113e-04 0.023648264 1  
## 68 control had 5 4.6677585 1.558113e-04 0.023648264 1  
## 69 control could 5 4.6677585 1.558113e-04 0.023648264 1  
## 70 control years 4 3.7342068 1.245945e-04 0.018918611 1  
## 71 control where 4 3.7342068 1.245945e-04 0.018918611 1  
## 72 control way 4 3.7342068 1.245945e-04 0.018918611 1  
## 73 control things 4 3.7342068 1.245945e-04 0.018918611 1  
## 74 control me 4 3.7342068 1.245945e-04 0.018918611 1  
## 75 control ireland 4 3.7342068 1.245945e-04 0.018918611 1  
## 76 control during 4 3.7342068 1.245945e-04 0.018918611 1  
## 77 control at 4 3.7342068 1.245945e-04 0.018918611 1  
## 78 control about 4 3.7342068 1.245945e-04 0.018918611 1  
## 79 control work 3 2.8006551 9.340499e-05 0.014188959 1  
## 80 control who 3 2.8006551 9.340499e-05 0.014188959 1  
## 81 control well 3 2.8006551 9.340499e-05 0.014188959 1  
## 82 control us 3 2.8006551 9.340499e-05 0.014188959 1  
## 83 control up 3 2.8006551 9.340499e-05 0.014188959 1  
## 84 control through 3 2.8006551 9.340499e-05 0.014188959 1  
## 85 control three 3 2.8006551 9.340499e-05 0.014188959 1  
## 86 control these 3 2.8006551 9.340499e-05 0.014188959 1  
## 87 control then 3 2.8006551 9.340499e-05 0.014188959 1  
## 88 control something 3 2.8006551 9.340499e-05 0.014188959 1  
## 89 control she 3 2.8006551 9.340499e-05 0.014188959 1  
## 90 control see 3 2.8006551 9.340499e-05 0.014188959 1  
## 91 control rt 3 2.8006551 9.340499e-05 0.014188959 1  
## 92 control race 3 2.8006551 9.340499e-05 0.014188959 1  
## 93 control probably 3 2.8006551 9.340499e-05 0.014188959 1  
## 94 control people 3 2.8006551 9.340499e-05 0.014188959 1  
## 95 control out 3 2.8006551 9.340499e-05 0.014188959 1  
## 96 control our 3 2.8006551 9.340499e-05 0.014188959 1  
## 97 control october 3 2.8006551 9.340499e-05 0.014188959 1  
## 98 control much 3 2.8006551 9.340499e-05 0.014188959 1  
## 99 control most 3 2.8006551 9.340499e-05 0.014188959 1  
## 100 control like 3 2.8006551 9.340499e-05 0.014188959 1  
## 101 control life 3 2.8006551 9.340499e-05 0.014188959 1  
## 102 control irish 3 2.8006551 9.340499e-05 0.014188959 1  
## 103 control http 3 2.8006551 9.340499e-05 0.014188959 1  
## 104 control how 3 2.8006551 9.340499e-05 0.014188959 1  
## 105 control her 3 2.8006551 9.340499e-05 0.014188959 1  
## 106 control great 3 2.8006551 9.340499e-05 0.014188959 1  
## 107 control government 3 2.8006551 9.340499e-05 0.014188959 1  
## 108 control going 3 2.8006551 9.340499e-05 0.014188959 1  
## 109 control day 3 2.8006551 9.340499e-05 0.014188959 1  
## 110 control between 3 2.8006551 9.340499e-05 0.014188959 1  
## 111 control already 3 2.8006551 9.340499e-05 0.014188959 1  
## 112 control year 2 1.8671034 6.224277e-05 0.009459306 1  
## 113 control world 2 1.8671034 6.224277e-05 0.009459306 1  
## 114 control working 2 1.8671034 6.224277e-05 0.009459306 1  
## 115 control without 2 1.8671034 6.224277e-05 0.009459306 1  
## 116 control went 2 1.8671034 6.224277e-05 0.009459306 1  
## 117 control various 2 1.8671034 6.224277e-05 0.009459306 1  
## 118 control used 2 1.8671034 6.224277e-05 0.009459306 1  
## 119 control university 2 1.8671034 6.224277e-05 0.009459306 1  
## 120 control two 2 1.8671034 6.224277e-05 0.009459306 1  
## 121 control too 2 1.8671034 6.224277e-05 0.009459306 1  
## 122 control time 2 1.8671034 6.224277e-05 0.009459306 1  
## 123 control those 2 1.8671034 6.224277e-05 0.009459306 1  
## 124 control them 2 1.8671034 6.224277e-05 0.009459306 1  
## 125 control teaching 2 1.8671034 6.224277e-05 0.009459306 1  
## 126 control still 2 1.8671034 6.224277e-05 0.009459306 1  
## 127 control state 2 1.8671034 6.224277e-05 0.009459306 1  
## 128 control started 2 1.8671034 6.224277e-05 0.009459306 1  
## 129 control space 2 1.8671034 6.224277e-05 0.009459306 1  
## 130 control somewhere 2 1.8671034 6.224277e-05 0.009459306 1  
## 131 control since 2 1.8671034 6.224277e-05 0.009459306 1  
## 132 control short 2 1.8671034 6.224277e-05 0.009459306 1  
## 133 control selling 2 1.8671034 6.224277e-05 0.009459306 1  
## 134 control says 2 1.8671034 6.224277e-05 0.009459306 1  
## 135 control same 2 1.8671034 6.224277e-05 0.009459306 1  
## 136 control risk 2 1.8671034 6.224277e-05 0.009459306 1  
## 137 control right 2 1.8671034 6.224277e-05 0.009459306 1  
## 138 control rest 2 1.8671034 6.224277e-05 0.009459306 1  
## 139 control research 2 1.8671034 6.224277e-05 0.009459306 1  
## 140 control require 2 1.8671034 6.224277e-05 0.009459306 1  
## 141 control politicians 2 1.8671034 6.224277e-05 0.009459306 1  
## 142 control performance 2 1.8671034 6.224277e-05 0.009459306 1  
## 143 control per 2 1.8671034 6.224277e-05 0.009459306 1  
## 144 control parts 2 1.8671034 6.224277e-05 0.009459306 1  
## 145 control own 2 1.8671034 6.224277e-05 0.009459306 1  
## 146 control ovens 2 1.8671034 6.224277e-05 0.009459306 1  
## 147 control officer 2 1.8671034 6.224277e-05 0.009459306 1  
## 148 control now 2 1.8671034 6.224277e-05 0.009459306 1  
## 149 control nice 2 1.8671034 6.224277e-05 0.009459306 1  
## 150 control nearly 2 1.8671034 6.224277e-05 0.009459306 1  
## 151 control more 2 1.8671034 6.224277e-05 0.009459306 1  
## 152 control mentioned 2 1.8671034 6.224277e-05 0.009459306 1  
## 153 control means 2 1.8671034 6.224277e-05 0.009459306 1  
## 154 control marriage 2 1.8671034 6.224277e-05 0.009459306 1  
## 155 control make 2 1.8671034 6.224277e-05 0.009459306 1  
## 156 control love 2 1.8671034 6.224277e-05 0.009459306 1  
## 157 control lot 2 1.8671034 6.224277e-05 0.009459306 1  
## 158 control looking 2 1.8671034 6.224277e-05 0.009459306 1  
## 159 control location 2 1.8671034 6.224277e-05 0.009459306 1  
## 160 control lived 2 1.8671034 6.224277e-05 0.009459306 1  
## 161 control laid 2 1.8671034 6.224277e-05 0.009459306 1  
## 162 control know 2 1.8671034 6.224277e-05 0.009459306 1  
## 163 control into 2 1.8671034 6.224277e-05 0.009459306 1  
## 164 control inclusion 2 1.8671034 6.224277e-05 0.009459306 1  
## 165 control however 2 1.8671034 6.224277e-05 0.009459306 1  
## 166 control house 2 1.8671034 6.224277e-05 0.009459306 1  
## 167 control here 2 1.8671034 6.224277e-05 0.009459306 1  
## 168 control hair 2 1.8671034 6.224277e-05 0.009459306 1  
## 169 control go 2 1.8671034 6.224277e-05 0.009459306 1  
## 170 control global 2 1.8671034 6.224277e-05 0.009459306 1  
## 171 control force 2 1.8671034 6.224277e-05 0.009459306 1  
## 172 control first 2 1.8671034 6.224277e-05 0.009459306 1  
## 173 control final 2 1.8671034 6.224277e-05 0.009459306 1  
## 174 control face 2 1.8671034 6.224277e-05 0.009459306 1  
## 175 control experience 2 1.8671034 6.224277e-05 0.009459306 1  
## 176 control evidence 2 1.8671034 6.224277e-05 0.009459306 1  
## 177 control even 2 1.8671034 6.224277e-05 0.009459306 1  
## 178 control each 2 1.8671034 6.224277e-05 0.009459306 1  
## 179 control does 2 1.8671034 6.224277e-05 0.009459306 1  
## 180 control courses 2 1.8671034 6.224277e-05 0.009459306 1  
## 181 control considered 2 1.8671034 6.224277e-05 0.009459306 1  
## 182 control care 2 1.8671034 6.224277e-05 0.009459306 1  
## 183 control carbon 2 1.8671034 6.224277e-05 0.009459306 1  
## 184 control can 2 1.8671034 6.224277e-05 0.009459306 1  
## 185 control called 2 1.8671034 6.224277e-05 0.009459306 1  
## 186 control business 2 1.8671034 6.224277e-05 0.009459306 1  
## 187 control broadcast 2 1.8671034 6.224277e-05 0.009459306 1  
## 188 control book 2 1.8671034 6.224277e-05 0.009459306 1  
## 189 control being 2 1.8671034 6.224277e-05 0.009459306 1  
## 190 control before 2 1.8671034 6.224277e-05 0.009459306 1  
## 191 control because 2 1.8671034 6.224277e-05 0.009459306 1  
## 192 control avoid 2 1.8671034 6.224277e-05 0.009459306 1  
## 193 control australia 2 1.8671034 6.224277e-05 0.009459306 1  
## 194 control around 2 1.8671034 6.224277e-05 0.009459306 1  
## 195 control another 2 1.8671034 6.224277e-05 0.009459306 1  
## 196 control also 2 1.8671034 6.224277e-05 0.009459306 1  
## 197 control after 2 1.8671034 6.224277e-05 0.009459306 1  
## 198 control 32 2 1.8671034 6.224277e-05 0.009459306 1  
## 199 control 2011 2 1.8671034 6.224277e-05 0.009459306 1  
## 200 control 20 2 1.8671034 6.224277e-05 0.009459306 1  
## 201 control 1 2 1.8671034 6.224277e-05 0.009459306 1  
## sig.chisq z p.z sig.z  
## 1 FALSE 8.87099005 0.000000e+00 TRUE  
## 2 FALSE 7.24313304 2.192690e-13 TRUE  
## 3 FALSE 7.24313304 2.192690e-13 TRUE  
## 4 FALSE 6.27273722 1.773777e-10 TRUE  
## 5 FALSE 5.69176518 6.286635e-09 TRUE  
## 6 FALSE 5.12166849 1.514220e-07 TRUE  
## 7 FALSE 5.12166849 1.514220e-07 TRUE  
## 8 FALSE 5.12166849 1.514220e-07 TRUE  
## 9 FALSE 5.12166849 1.514220e-07 TRUE  
## 10 FALSE 5.12166849 1.514220e-07 TRUE  
## 11 FALSE 5.12166849 1.514220e-07 TRUE  
## 12 FALSE 5.12166849 1.514220e-07 TRUE  
## 13 FALSE 5.12166849 1.514220e-07 TRUE  
## 14 FALSE 5.12166849 1.514220e-07 TRUE  
## 15 FALSE 5.12166849 1.514220e-07 TRUE  
## 16 FALSE 1.51851713 6.444204e-02 FALSE  
## 17 FALSE 0.81951858 2.062453e-01 FALSE  
## 18 FALSE 0.54586457 2.925795e-01 FALSE  
## 19 FALSE 0.52375554 3.002243e-01 FALSE  
## 20 FALSE 0.47646966 3.168699e-01 FALSE  
## 21 FALSE 0.45097126 3.260051e-01 FALSE  
## 22 FALSE 0.40100897 3.442068e-01 FALSE  
## 23 FALSE 0.38290892 3.508936e-01 FALSE  
## 24 FALSE 0.36390971 3.579627e-01 FALSE  
## 25 FALSE 0.34386236 3.654749e-01 FALSE  
## 26 FALSE 0.33691493 3.680905e-01 FALSE  
## 27 FALSE 0.29177689 3.852286e-01 FALSE  
## 28 FALSE 0.28355616 3.883753e-01 FALSE  
## 29 FALSE 0.27508989 3.916236e-01 FALSE  
## 30 FALSE 0.27508989 3.916236e-01 FALSE  
## 31 FALSE 0.25732303 3.984647e-01 FALSE  
## 32 FALSE 0.25732303 3.984647e-01 FALSE  
## 33 FALSE 0.25732303 3.984647e-01 FALSE  
## 34 FALSE 0.25732303 3.984647e-01 FALSE  
## 35 FALSE 0.25732303 3.984647e-01 FALSE  
## 36 FALSE 0.23823483 4.058495e-01 FALSE  
## 37 FALSE 0.23823483 4.058495e-01 FALSE  
## 38 FALSE 0.22809248 4.097872e-01 FALSE  
## 39 FALSE 0.21747765 4.139181e-01 FALSE  
## 40 FALSE 0.21747765 4.139181e-01 FALSE  
## 41 FALSE 0.20631742 4.182715e-01 FALSE  
## 42 FALSE 0.20631742 4.182715e-01 FALSE  
## 43 FALSE 0.20631742 4.182715e-01 FALSE  
## 44 FALSE 0.20631742 4.182715e-01 FALSE  
## 45 FALSE 0.19451792 4.228852e-01 FALSE  
## 46 FALSE 0.19451792 4.228852e-01 FALSE  
## 47 FALSE 0.18195486 4.278091e-01 FALSE  
## 48 FALSE 0.18195486 4.278091e-01 FALSE  
## 49 FALSE 0.18195486 4.278091e-01 FALSE  
## 50 FALSE 0.18195486 4.278091e-01 FALSE  
## 51 FALSE 0.18195486 4.278091e-01 FALSE  
## 52 FALSE 0.16845746 4.331117e-01 FALSE  
## 53 FALSE 0.16845746 4.331117e-01 FALSE  
## 54 FALSE 0.16845746 4.331117e-01 FALSE  
## 55 FALSE 0.16845746 4.331117e-01 FALSE  
## 56 FALSE 0.16845746 4.331117e-01 FALSE  
## 57 FALSE 0.16845746 4.331117e-01 FALSE  
## 58 FALSE 0.16845746 4.331117e-01 FALSE  
## 59 FALSE 0.15377992 4.388916e-01 FALSE  
## 60 FALSE 0.15377992 4.388916e-01 FALSE  
## 61 FALSE 0.15377992 4.388916e-01 FALSE  
## 62 FALSE 0.15377992 4.388916e-01 FALSE  
## 63 FALSE 0.15377992 4.388916e-01 FALSE  
## 64 FALSE 0.15377992 4.388916e-01 FALSE  
## 65 FALSE 0.15377992 4.388916e-01 FALSE  
## 66 FALSE 0.15377992 4.388916e-01 FALSE  
## 67 FALSE 0.15377992 4.388916e-01 FALSE  
## 68 FALSE 0.15377992 4.388916e-01 FALSE  
## 69 FALSE 0.15377992 4.388916e-01 FALSE  
## 70 FALSE 0.13754494 4.453000e-01 FALSE  
## 71 FALSE 0.13754494 4.453000e-01 FALSE  
## 72 FALSE 0.13754494 4.453000e-01 FALSE  
## 73 FALSE 0.13754494 4.453000e-01 FALSE  
## 74 FALSE 0.13754494 4.453000e-01 FALSE  
## 75 FALSE 0.13754494 4.453000e-01 FALSE  
## 76 FALSE 0.13754494 4.453000e-01 FALSE  
## 77 FALSE 0.13754494 4.453000e-01 FALSE  
## 78 FALSE 0.13754494 4.453000e-01 FALSE  
## 79 FALSE 0.11911742 4.525912e-01 FALSE  
## 80 FALSE 0.11911742 4.525912e-01 FALSE  
## 81 FALSE 0.11911742 4.525912e-01 FALSE  
## 82 FALSE 0.11911742 4.525912e-01 FALSE  
## 83 FALSE 0.11911742 4.525912e-01 FALSE  
## 84 FALSE 0.11911742 4.525912e-01 FALSE  
## 85 FALSE 0.11911742 4.525912e-01 FALSE  
## 86 FALSE 0.11911742 4.525912e-01 FALSE  
## 87 FALSE 0.11911742 4.525912e-01 FALSE  
## 88 FALSE 0.11911742 4.525912e-01 FALSE  
## 89 FALSE 0.11911742 4.525912e-01 FALSE  
## 90 FALSE 0.11911742 4.525912e-01 FALSE  
## 91 FALSE 0.11911742 4.525912e-01 FALSE  
## 92 FALSE 0.11911742 4.525912e-01 FALSE  
## 93 FALSE 0.11911742 4.525912e-01 FALSE  
## 94 FALSE 0.11911742 4.525912e-01 FALSE  
## 95 FALSE 0.11911742 4.525912e-01 FALSE  
## 96 FALSE 0.11911742 4.525912e-01 FALSE  
## 97 FALSE 0.11911742 4.525912e-01 FALSE  
## 98 FALSE 0.11911742 4.525912e-01 FALSE  
## 99 FALSE 0.11911742 4.525912e-01 FALSE  
## 100 FALSE 0.11911742 4.525912e-01 FALSE  
## 101 FALSE 0.11911742 4.525912e-01 FALSE  
## 102 FALSE 0.11911742 4.525912e-01 FALSE  
## 103 FALSE 0.11911742 4.525912e-01 FALSE  
## 104 FALSE 0.11911742 4.525912e-01 FALSE  
## 105 FALSE 0.11911742 4.525912e-01 FALSE  
## 106 FALSE 0.11911742 4.525912e-01 FALSE  
## 107 FALSE 0.11911742 4.525912e-01 FALSE  
## 108 FALSE 0.11911742 4.525912e-01 FALSE  
## 109 FALSE 0.11911742 4.525912e-01 FALSE  
## 110 FALSE 0.11911742 4.525912e-01 FALSE  
## 111 FALSE 0.11911742 4.525912e-01 FALSE  
## 112 FALSE 0.09725896 4.612604e-01 FALSE  
## 113 FALSE 0.09725896 4.612604e-01 FALSE  
## 114 FALSE 0.09725896 4.612604e-01 FALSE  
## 115 FALSE 0.09725896 4.612604e-01 FALSE  
## 116 FALSE 0.09725896 4.612604e-01 FALSE  
## 117 FALSE 0.09725896 4.612604e-01 FALSE  
## 118 FALSE 0.09725896 4.612604e-01 FALSE  
## 119 FALSE 0.09725896 4.612604e-01 FALSE  
## 120 FALSE 0.09725896 4.612604e-01 FALSE  
## 121 FALSE 0.09725896 4.612604e-01 FALSE  
## 122 FALSE 0.09725896 4.612604e-01 FALSE  
## 123 FALSE 0.09725896 4.612604e-01 FALSE  
## 124 FALSE 0.09725896 4.612604e-01 FALSE  
## 125 FALSE 0.09725896 4.612604e-01 FALSE  
## 126 FALSE 0.09725896 4.612604e-01 FALSE  
## 127 FALSE 0.09725896 4.612604e-01 FALSE  
## 128 FALSE 0.09725896 4.612604e-01 FALSE  
## 129 FALSE 0.09725896 4.612604e-01 FALSE  
## 130 FALSE 0.09725896 4.612604e-01 FALSE  
## 131 FALSE 0.09725896 4.612604e-01 FALSE  
## 132 FALSE 0.09725896 4.612604e-01 FALSE  
## 133 FALSE 0.09725896 4.612604e-01 FALSE  
## 134 FALSE 0.09725896 4.612604e-01 FALSE  
## 135 FALSE 0.09725896 4.612604e-01 FALSE  
## 136 FALSE 0.09725896 4.612604e-01 FALSE  
## 137 FALSE 0.09725896 4.612604e-01 FALSE  
## 138 FALSE 0.09725896 4.612604e-01 FALSE  
## 139 FALSE 0.09725896 4.612604e-01 FALSE  
## 140 FALSE 0.09725896 4.612604e-01 FALSE  
## 141 FALSE 0.09725896 4.612604e-01 FALSE  
## 142 FALSE 0.09725896 4.612604e-01 FALSE  
## 143 FALSE 0.09725896 4.612604e-01 FALSE  
## 144 FALSE 0.09725896 4.612604e-01 FALSE  
## 145 FALSE 0.09725896 4.612604e-01 FALSE  
## 146 FALSE 0.09725896 4.612604e-01 FALSE  
## 147 FALSE 0.09725896 4.612604e-01 FALSE  
## 148 FALSE 0.09725896 4.612604e-01 FALSE  
## 149 FALSE 0.09725896 4.612604e-01 FALSE  
## 150 FALSE 0.09725896 4.612604e-01 FALSE  
## 151 FALSE 0.09725896 4.612604e-01 FALSE  
## 152 FALSE 0.09725896 4.612604e-01 FALSE  
## 153 FALSE 0.09725896 4.612604e-01 FALSE  
## 154 FALSE 0.09725896 4.612604e-01 FALSE  
## 155 FALSE 0.09725896 4.612604e-01 FALSE  
## 156 FALSE 0.09725896 4.612604e-01 FALSE  
## 157 FALSE 0.09725896 4.612604e-01 FALSE  
## 158 FALSE 0.09725896 4.612604e-01 FALSE  
## 159 FALSE 0.09725896 4.612604e-01 FALSE  
## 160 FALSE 0.09725896 4.612604e-01 FALSE  
## 161 FALSE 0.09725896 4.612604e-01 FALSE  
## 162 FALSE 0.09725896 4.612604e-01 FALSE  
## 163 FALSE 0.09725896 4.612604e-01 FALSE  
## 164 FALSE 0.09725896 4.612604e-01 FALSE  
## 165 FALSE 0.09725896 4.612604e-01 FALSE  
## 166 FALSE 0.09725896 4.612604e-01 FALSE  
## 167 FALSE 0.09725896 4.612604e-01 FALSE  
## 168 FALSE 0.09725896 4.612604e-01 FALSE  
## 169 FALSE 0.09725896 4.612604e-01 FALSE  
## 170 FALSE 0.09725896 4.612604e-01 FALSE  
## 171 FALSE 0.09725896 4.612604e-01 FALSE  
## 172 FALSE 0.09725896 4.612604e-01 FALSE  
## 173 FALSE 0.09725896 4.612604e-01 FALSE  
## 174 FALSE 0.09725896 4.612604e-01 FALSE  
## 175 FALSE 0.09725896 4.612604e-01 FALSE  
## 176 FALSE 0.09725896 4.612604e-01 FALSE  
## 177 FALSE 0.09725896 4.612604e-01 FALSE  
## 178 FALSE 0.09725896 4.612604e-01 FALSE  
## 179 FALSE 0.09725896 4.612604e-01 FALSE  
## 180 FALSE 0.09725896 4.612604e-01 FALSE  
## 181 FALSE 0.09725896 4.612604e-01 FALSE  
## 182 FALSE 0.09725896 4.612604e-01 FALSE  
## 183 FALSE 0.09725896 4.612604e-01 FALSE  
## 184 FALSE 0.09725896 4.612604e-01 FALSE  
## 185 FALSE 0.09725896 4.612604e-01 FALSE  
## 186 FALSE 0.09725896 4.612604e-01 FALSE  
## 187 FALSE 0.09725896 4.612604e-01 FALSE  
## 188 FALSE 0.09725896 4.612604e-01 FALSE  
## 189 FALSE 0.09725896 4.612604e-01 FALSE  
## 190 FALSE 0.09725896 4.612604e-01 FALSE  
## 191 FALSE 0.09725896 4.612604e-01 FALSE  
## 192 FALSE 0.09725896 4.612604e-01 FALSE  
## 193 FALSE 0.09725896 4.612604e-01 FALSE  
## 194 FALSE 0.09725896 4.612604e-01 FALSE  
## 195 FALSE 0.09725896 4.612604e-01 FALSE  
## 196 FALSE 0.09725896 4.612604e-01 FALSE  
## 197 FALSE 0.09725896 4.612604e-01 FALSE  
## 198 FALSE 0.09725896 4.612604e-01 FALSE  
## 199 FALSE 0.09725896 4.612604e-01 FALSE  
## 200 FALSE 0.09725896 4.612604e-01 FALSE  
## 201 FALSE 0.09725896 4.612604e-01 FALSE  
##   
##   
## Summary statistics:  
##   
## Total Chi squared = 525.5987   
## Total degrees of freedom = 199   
## p = 0   
## Sum of counts = 2137   
##   
## Levels:  
##   
## control x   
## 2 200

r7 <- cfa::cfa(configs, counts)  
# save to disc  
sig\_words <- r7$table %>%  
 as.data.frame() %>%  
 dplyr::mutate(expected = round(expected, 1),  
 Q = round(Q, 3),  
 chisq = round(chisq, 3),  
 z = round(z, 3),  
 p.z = round(p.z, 3),  
 label = stringr::str\_remove\_all(label, "test ")) %>%  
 dplyr::select(-p.chisq, -sig.chisq)  
# save to disc  
write.table(sig\_words,   
 here::here("tables", "xmuch\_words.txt"),   
 sep = "\t",   
 row.names = F)  
# extract words  
words <- r7$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("test ")  
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(control == "test") %>%  
 dplyr::group\_by(variety, x) %>%  
 dplyr::mutate(freq = n()) %>%  
 dplyr::mutate(x = ifelse(freq < 2, "other", x)) %>%  
 dplyr::ungroup() %>%  
 dplyr::group\_by(x, variety) %>%  
 dplyr::summarise(freq = n())

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

configs <- wordcat\_tb %>%  
 dplyr::select(variety, x)  
counts = wordcat\_tb$freq  
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 6 4.6056338 0.010148642 0.422147564 0.9999971  
## 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 26.4929577 0.004389709 0.009704158 1.0000000  
## 13 Australia other 87 87.5070423 0.005794302 0.002937956 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 0.64972884 0.25793370 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 0.09850968 0.46076379 FALSE  
## 13 FALSE 0.05420292 0.47838675 FALSE  
##   
##   
## Summary statistics:  
##   
## Total Chi squared = 13.16728   
## Total degrees of freedom = 10   
## p = 0.2144687   
## 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("test ")  
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(control, variety, x) %>%  
 dplyr::mutate(freq = n()) %>%  
 dplyr::mutate(x = ifelse(freq < 2, "other", x)) %>%  
 dplyr::ungroup() %>%  
 dplyr::group\_by(x, variety, control) %>%  
 dplyr::summarise(freq = n())

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

configs <- wordcat\_tb %>%  
 dplyr::select(control, variety, x)  
counts = wordcat\_tb$freq  
cfa::cfa(configs, counts)

##   
## \*\*\* Analysis of configuration frequencies (CFA) \*\*\*  
##   
## label n expected Q chisq  
## 1 test Australia jealous 6 0.20596170 2.710288e-03 1.629957e+02  
## 2 test Australia threatened 3 0.10298085 1.355079e-03 8.149787e+01  
## 3 test Australia paranoid 3 0.10298085 1.355079e-03 8.149787e+01  
## 4 test Australia other 87 35.45973961 2.451333e-02 7.491308e+01  
## 5 test Ireland excuses 2 0.06418052 9.054619e-04 5.838838e+01  
## 6 test Ireland coincidence 2 0.06418052 9.054619e-04 5.838838e+01  
## 7 test Australia hypocrite 2 0.06865390 9.033714e-04 5.433191e+01  
## 8 test Australia generalise 2 0.06865390 9.033714e-04 5.433191e+01  
## 9 test Australia entitled 2 0.06865390 9.033714e-04 5.433191e+01  
## 10 test Australia creepy 2 0.06865390 9.033714e-04 5.433191e+01  
## 11 test Ireland agenda 2 0.12836105 8.754683e-04 2.729046e+01  
## 12 test Australia agenda 2 0.13730780 8.712871e-04 2.526894e+01  
## 13 control Australia other 436 498.43408639 3.807964e-02 7.820523e+00  
## 14 control Ireland this 14 7.66821671 2.972205e-03 5.228266e+00  
## 15 control Australia would 8 3.86008973 1.939850e-03 4.440015e+00  
## 16 control Australia not 8 3.86008973 1.939850e-03 4.440015e+00  
## 17 control Ireland radio 5 2.25535786 1.285098e-03 3.340073e+00  
## 18 control Ireland could 5 2.25535786 1.285098e-03 3.340073e+00  
## 19 control Australia as 6 2.89506730 1.454230e-03 3.330011e+00  
## 20 control Australia this 3 8.20269068 2.442810e-03 3.299892e+00  
## 21 control Australia a 29 20.74798230 3.897513e-03 3.282044e+00  
## 22 control Australia which 5 2.41255608 1.211584e-03 2.775009e+00  
## 23 control Ireland years 4 1.80428629 1.027862e-03 2.672059e+00  
## 24 control Ireland ireland 4 1.80428629 1.027862e-03 2.672059e+00  
## 25 control Australia had 4 1.93004487 9.690484e-04 2.220008e+00  
## 26 control Ireland way 3 1.35321471 7.707335e-04 2.004044e+00  
## 27 control Ireland three 3 1.35321471 7.707335e-04 2.004044e+00  
## 28 control Ireland something 3 1.35321471 7.707335e-04 2.004044e+00  
## 29 control Ireland see 3 1.35321471 7.707335e-04 2.004044e+00  
## 30 control Ireland rt 3 1.35321471 7.707335e-04 2.004044e+00  
## 31 control Ireland race 3 1.35321471 7.707335e-04 2.004044e+00  
## 32 control Ireland probably 3 1.35321471 7.707335e-04 2.004044e+00  
## 33 control Ireland our 3 1.35321471 7.707335e-04 2.004044e+00  
## 34 control Ireland me 3 1.35321471 7.707335e-04 2.004044e+00  
## 35 control Ireland http 3 1.35321471 7.707335e-04 2.004044e+00  
## 36 control Ireland her 3 1.35321471 7.707335e-04 2.004044e+00  
## 37 control Ireland day 3 1.35321471 7.707335e-04 2.004044e+00  
## 38 control Ireland the 75 64.05216314 5.278743e-03 1.871211e+00  
## 39 control Australia things 3 1.44753365 7.266222e-04 1.665006e+00  
## 40 control Australia people 3 1.44753365 7.266222e-04 1.665006e+00  
## 41 control Australia like 3 1.44753365 7.266222e-04 1.665006e+00  
## 42 control Australia during 3 1.44753365 7.266222e-04 1.665006e+00  
## 43 control Australia between 3 1.44753365 7.266222e-04 1.665006e+00  
## 44 control Australia about 3 1.44753365 7.266222e-04 1.665006e+00  
## 45 control Australia we 7 4.34260095 1.245467e-03 1.626161e+00  
## 46 control Australia have 10 6.75515703 1.522511e-03 1.558662e+00  
## 47 control Ireland a 14 19.39607757 2.546997e-03 1.501213e+00  
## 48 control Australia for 11 7.72017946 1.539620e-03 1.393390e+00  
## 49 control Ireland year 2 0.90214314 5.137139e-04 1.336029e+00  
## 50 control Ireland who 2 0.90214314 5.137139e-04 1.336029e+00  
## 51 control Ireland various 2 0.90214314 5.137139e-04 1.336029e+00  
## 52 control Ireland then 2 0.90214314 5.137139e-04 1.336029e+00  
## 53 control Ireland still 2 0.90214314 5.137139e-04 1.336029e+00  
## 54 control Ireland short 2 0.90214314 5.137139e-04 1.336029e+00  
## 55 control Ireland selling 2 0.90214314 5.137139e-04 1.336029e+00  
## 56 control Ireland october 2 0.90214314 5.137139e-04 1.336029e+00  
## 57 control Ireland much 2 0.90214314 5.137139e-04 1.336029e+00  
## 58 control Ireland most 2 0.90214314 5.137139e-04 1.336029e+00  
## 59 control Ireland make 2 0.90214314 5.137139e-04 1.336029e+00  
## 60 control Ireland looking 2 0.90214314 5.137139e-04 1.336029e+00  
## 61 control Ireland life 2 0.90214314 5.137139e-04 1.336029e+00  
## 62 control Ireland laid 2 0.90214314 5.137139e-04 1.336029e+00  
## 63 control Ireland irish 2 0.90214314 5.137139e-04 1.336029e+00  
## 64 control Ireland how 2 0.90214314 5.137139e-04 1.336029e+00  
## 65 control Ireland government 2 0.90214314 5.137139e-04 1.336029e+00  
## 66 control Ireland going 2 0.90214314 5.137139e-04 1.336029e+00  
## 67 control Ireland first 2 0.90214314 5.137139e-04 1.336029e+00  
## 68 control Ireland face 2 0.90214314 5.137139e-04 1.336029e+00  
## 69 control Ireland broadcast 2 0.90214314 5.137139e-04 1.336029e+00  
## 70 control Ireland being 2 0.90214314 5.137139e-04 1.336029e+00  
## 71 control Ireland around 2 0.90214314 5.137139e-04 1.336029e+00  
## 72 control Ireland 1 2 0.90214314 5.137139e-04 1.336029e+00  
## 73 control Ireland and 32 26.16215114 2.764345e-03 1.302663e+00  
## 74 control Ireland be 8 5.41285886 1.213147e-03 1.236555e+00  
## 75 control Ireland is 18 13.98321871 1.891125e-03 1.153850e+00  
## 76 control Ireland by 9 6.31500200 1.259566e-03 1.141601e+00  
## 77 test Ireland other 27 33.14924074 2.921462e-03 1.140695e+00  
## 78 control Australia work 2 0.96502243 4.843054e-04 1.110004e+00  
## 79 control Australia went 2 0.96502243 4.843054e-04 1.110004e+00  
## 80 control Australia well 2 0.96502243 4.843054e-04 1.110004e+00  
## 81 control Australia used 2 0.96502243 4.843054e-04 1.110004e+00  
## 82 control Australia us 2 0.96502243 4.843054e-04 1.110004e+00  
## 83 control Australia up 2 0.96502243 4.843054e-04 1.110004e+00  
## 84 control Australia university 2 0.96502243 4.843054e-04 1.110004e+00  
## 85 control Australia time 2 0.96502243 4.843054e-04 1.110004e+00  
## 86 control Australia through 2 0.96502243 4.843054e-04 1.110004e+00  
## 87 control Australia these 2 0.96502243 4.843054e-04 1.110004e+00  
## 88 control Australia teaching 2 0.96502243 4.843054e-04 1.110004e+00  
## 89 control Australia started 2 0.96502243 4.843054e-04 1.110004e+00  
## 90 control Australia somewhere 2 0.96502243 4.843054e-04 1.110004e+00  
## 91 control Australia since 2 0.96502243 4.843054e-04 1.110004e+00  
## 92 control Australia she 2 0.96502243 4.843054e-04 1.110004e+00  
## 93 control Australia says 2 0.96502243 4.843054e-04 1.110004e+00  
## 94 control Australia right 2 0.96502243 4.843054e-04 1.110004e+00  
## 95 control Australia rest 2 0.96502243 4.843054e-04 1.110004e+00  
## 96 control Australia parts 2 0.96502243 4.843054e-04 1.110004e+00  
## 97 control Australia own 2 0.96502243 4.843054e-04 1.110004e+00  
## 98 control Australia ovens 2 0.96502243 4.843054e-04 1.110004e+00  
## 99 control Australia out 2 0.96502243 4.843054e-04 1.110004e+00  
## 100 control Australia officer 2 0.96502243 4.843054e-04 1.110004e+00  
## 101 control Australia now 2 0.96502243 4.843054e-04 1.110004e+00  
## 102 control Australia means 2 0.96502243 4.843054e-04 1.110004e+00  
## 103 control Australia marriage 2 0.96502243 4.843054e-04 1.110004e+00  
## 104 control Australia inclusion 2 0.96502243 4.843054e-04 1.110004e+00  
## 105 control Australia here 2 0.96502243 4.843054e-04 1.110004e+00  
## 106 control Australia hair 2 0.96502243 4.843054e-04 1.110004e+00  
## 107 control Australia great 2 0.96502243 4.843054e-04 1.110004e+00  
## 108 control Australia go 2 0.96502243 4.843054e-04 1.110004e+00  
## 109 control Australia global 2 0.96502243 4.843054e-04 1.110004e+00  
## 110 control Australia final 2 0.96502243 4.843054e-04 1.110004e+00  
## 111 control Australia experience 2 0.96502243 4.843054e-04 1.110004e+00  
## 112 control Australia even 2 0.96502243 4.843054e-04 1.110004e+00  
## 113 control Australia courses 2 0.96502243 4.843054e-04 1.110004e+00  
## 114 control Australia carbon 2 0.96502243 4.843054e-04 1.110004e+00  
## 115 control Australia can 2 0.96502243 4.843054e-04 1.110004e+00  
## 116 control Australia before 2 0.96502243 4.843054e-04 1.110004e+00  
## 117 control Australia australia 2 0.96502243 4.843054e-04 1.110004e+00  
## 118 control Australia another 2 0.96502243 4.843054e-04 1.110004e+00  
## 119 control Australia also 2 0.96502243 4.843054e-04 1.110004e+00  
## 120 control Australia already 2 0.96502243 4.843054e-04 1.110004e+00  
## 121 control Australia after 2 0.96502243 4.843054e-04 1.110004e+00  
## 122 control Australia 2011 2 0.96502243 4.843054e-04 1.110004e+00  
## 123 control Australia 20 2 0.96502243 4.843054e-04 1.110004e+00  
## 124 control Ireland when 5 3.15750100 8.630609e-04 1.075155e+00  
## 125 control Ireland but 5 3.15750100 8.630609e-04 1.075155e+00  
## 126 control Ireland we 2 4.05964414 9.651836e-04 1.044952e+00  
## 127 control Australia of 28 23.16053838 2.288335e-03 1.011220e+00  
## 128 control Ireland have 4 6.31500200 1.085996e-03 8.486512e-01  
## 129 control Australia if 8 5.79013460 1.036420e-03 8.434182e-01  
## 130 control Ireland an 7 4.96178729 9.555444e-04 8.372610e-01  
## 131 control Ireland to 33 28.41750900 2.172227e-03 7.389537e-01  
## 132 control Ireland for 5 7.21714514 1.040531e-03 6.811187e-01  
## 133 control Ireland i 14 11.27678928 1.280473e-03 6.576231e-01  
## 134 control Australia from 6 4.34260095 7.767878e-04 6.325637e-01  
## 135 control Ireland other 483 465.95693325 1.019296e-02 6.233755e-01  
## 136 control Ireland their 4 2.70642943 6.058046e-04 6.182776e-01  
## 137 control Ireland do 4 2.70642943 6.058046e-04 6.182776e-01  
## 138 control Australia when 2 3.37757851 6.453500e-04 5.618589e-01  
## 139 control Australia but 2 3.37757851 6.453500e-04 5.618589e-01  
## 140 control Australia be 4 5.79013460 8.395677e-04 5.534555e-01  
## 141 control Ireland were 5 3.60857257 6.519083e-04 5.365197e-01  
## 142 control Ireland they 6 4.51071571 6.980510e-04 4.917108e-01  
## 143 control Ireland in 18 15.33643343 1.254823e-03 4.625969e-01  
## 144 control Australia by 5 6.75515703 8.235361e-04 4.560332e-01  
## 145 control Australia your 4 2.89506730 5.175074e-04 4.217091e-01  
## 146 control Australia what 4 2.89506730 5.175074e-04 4.217091e-01  
## 147 control Australia there 4 2.89506730 5.175074e-04 4.217091e-01  
## 148 control Ireland if 4 5.41285886 6.625093e-04 3.687830e-01  
## 149 control Australia will 5 3.86008973 5.341310e-04 3.366231e-01  
## 150 control Australia an 4 5.30762338 6.131327e-04 3.221553e-01  
## 151 control Australia their 2 2.89506730 4.192147e-04 2.767278e-01  
## 152 control Australia do 2 2.89506730 4.192147e-04 2.767278e-01  
## 153 control Ireland from 3 4.05964414 4.965669e-04 2.765872e-01  
## 154 control Australia is 13 14.95784770 9.221897e-04 2.562647e-01  
## 155 control Ireland only 3 2.25535786 3.486569e-04 2.458554e-01  
## 156 control Ireland its 3 2.25535786 3.486569e-04 2.458554e-01  
## 157 control Australia are 8 6.75515703 5.840920e-04 2.294001e-01  
## 158 control Australia were 3 3.86008973 4.030147e-04 1.916417e-01  
## 159 control Ireland your 2 2.70642943 3.308348e-04 1.843915e-01  
## 160 control Ireland what 2 2.70642943 3.308348e-04 1.843915e-01  
## 161 control Ireland there 2 2.70642943 3.308348e-04 1.843915e-01  
## 162 control Australia it 13 11.58026919 6.676625e-04 1.740577e-01  
## 163 control Australia that 15 13.51031406 7.011971e-04 1.642570e-01  
## 164 control Australia than 3 2.41255608 2.750737e-04 1.430393e-01  
## 165 control Australia so 3 2.41255608 2.750737e-04 1.430393e-01  
## 166 control Australia over 3 2.41255608 2.750737e-04 1.430393e-01  
## 167 control Australia my 3 2.41255608 2.750737e-04 1.430393e-01  
## 168 control Australia just 3 2.41255608 2.750737e-04 1.430393e-01  
## 169 control Australia has 3 2.41255608 2.750737e-04 1.430393e-01  
## 170 control Australia they 4 4.82511216 3.868001e-04 1.410973e-01  
## 171 control Australia and 26 27.98565054 9.410602e-04 1.408868e-01  
## 172 control Ireland of 20 21.65143543 7.803230e-04 1.259611e-01  
## 173 control Australia one 4 3.37757851 2.915839e-04 1.147001e-01  
## 174 control Australia all 4 3.37757851 2.915839e-04 1.147001e-01  
## 175 control Ireland will 3 3.60857257 2.851270e-04 1.026335e-01  
## 176 control Ireland with 9 8.11928829 4.135028e-04 9.553216e-02  
## 177 control Australia i 11 12.06278041 4.999115e-04 9.363531e-02  
## 178 control Ireland on 8 7.21714514 3.674025e-04 8.491747e-02  
## 179 control Ireland you 7 6.31500200 3.213411e-04 7.430279e-02  
## 180 control Ireland was 7 6.31500200 3.213411e-04 7.430279e-02  
## 181 control Australia only 2 2.41255608 1.931815e-04 7.054863e-02  
## 182 control Australia its 2 2.41255608 1.931815e-04 7.054863e-02  
## 183 control Ireland he 5 4.51071571 2.293352e-04 5.307342e-02  
## 184 control Australia the 67 68.51659271 7.328364e-04 3.356929e-02  
## 185 control Ireland or 3 2.70642943 1.374849e-04 3.184405e-02  
## 186 control Ireland than 2 2.25535786 1.195639e-04 2.891232e-02  
## 187 control Ireland so 2 2.25535786 1.195639e-04 2.891232e-02  
## 188 control Ireland over 2 2.25535786 1.195639e-04 2.891232e-02  
## 189 control Ireland my 2 2.25535786 1.195639e-04 2.891232e-02  
## 190 control Ireland just 2 2.25535786 1.195639e-04 2.891232e-02  
## 191 control Ireland has 2 2.25535786 1.195639e-04 2.891232e-02  
## 192 control Ireland where 2 1.80428629 9.161788e-05 2.122937e-02  
## 193 control Ireland at 2 1.80428629 9.161788e-05 2.122937e-02  
## 194 control Ireland are 6 6.31500200 1.477714e-04 1.571278e-02  
## 195 control Australia with 9 8.68520189 1.478401e-04 1.140996e-02  
## 196 control Ireland that 13 12.63000400 1.740855e-04 1.083903e-02  
## 197 control Australia on 8 7.72017946 1.313539e-04 1.014219e-02  
## 198 control Australia in 16 16.40538135 1.910739e-04 1.001708e-02  
## 199 control Australia you 7 6.75515703 1.148826e-04 8.874417e-03  
## 200 control Australia was 7 6.75515703 1.148826e-04 8.874417e-03  
## 201 control Ireland one 3 3.15750100 7.377640e-05 7.856392e-03  
## 202 control Ireland all 3 3.15750100 7.377640e-05 7.856392e-03  
## 203 control Australia he 5 4.82511216 8.198476e-05 6.338869e-03  
## 204 control Australia to 30 30.39820662 1.889383e-04 5.216377e-03  
## 205 control Australia or 3 2.89506730 4.914639e-05 3.803322e-03  
## 206 control Ireland it 11 10.82571771 8.193136e-05 2.805755e-03  
## 207 control Australia where 2 1.93004487 3.274946e-05 2.535548e-03  
## 208 control Australia at 2 1.93004487 3.274946e-05 2.535548e-03  
## p.chisq sig.chisq z p.z sig.z  
## 1 0.3141002 FALSE 12.76697847 0.000000e+00 TRUE  
## 2 0.9999998 FALSE 9.02761705 0.000000e+00 TRUE  
## 3 0.9999998 FALSE 9.02761705 0.000000e+00 TRUE  
## 4 1.0000000 FALSE 8.65523447 0.000000e+00 TRUE  
## 5 1.0000000 FALSE 7.64122917 1.076916e-14 TRUE  
## 6 1.0000000 FALSE 7.64122917 1.076916e-14 TRUE  
## 7 1.0000000 FALSE 7.37101846 8.471002e-14 TRUE  
## 8 1.0000000 FALSE 7.37101846 8.471002e-14 TRUE  
## 9 1.0000000 FALSE 7.37101846 8.471002e-14 TRUE  
## 10 1.0000000 FALSE 7.37101846 8.471002e-14 TRUE  
## 11 1.0000000 FALSE 5.22402741 8.753642e-08 TRUE  
## 12 1.0000000 FALSE 5.02682180 2.493377e-07 TRUE  
## 13 1.0000000 FALSE 2.79651977 2.582813e-03 FALSE  
## 14 1.0000000 FALSE 2.28654026 1.111134e-02 FALSE  
## 15 1.0000000 FALSE 2.10713436 1.755297e-02 FALSE  
## 16 1.0000000 FALSE 2.10713436 1.755297e-02 FALSE  
## 17 1.0000000 FALSE 1.82758673 3.380580e-02 FALSE  
## 18 1.0000000 FALSE 1.82758673 3.380580e-02 FALSE  
## 19 1.0000000 FALSE 1.82483188 3.401320e-02 FALSE  
## 20 1.0000000 FALSE 1.81656038 3.464223e-02 FALSE  
## 21 1.0000000 FALSE 1.81164132 3.502082e-02 FALSE  
## 22 1.0000000 FALSE 1.66583597 4.787304e-02 FALSE  
## 23 1.0000000 FALSE 1.63464327 5.106193e-02 FALSE  
## 24 1.0000000 FALSE 1.63464327 5.106193e-02 FALSE  
## 25 1.0000000 FALSE 1.48996899 6.811619e-02 FALSE  
## 26 1.0000000 FALSE 1.41564260 7.844009e-02 FALSE  
## 27 1.0000000 FALSE 1.41564260 7.844009e-02 FALSE  
## 28 1.0000000 FALSE 1.41564260 7.844009e-02 FALSE  
## 29 1.0000000 FALSE 1.41564260 7.844009e-02 FALSE  
## 30 1.0000000 FALSE 1.41564260 7.844009e-02 FALSE  
## 31 1.0000000 FALSE 1.41564260 7.844009e-02 FALSE  
## 32 1.0000000 FALSE 1.41564260 7.844009e-02 FALSE  
## 33 1.0000000 FALSE 1.41564260 7.844009e-02 FALSE  
## 34 1.0000000 FALSE 1.41564260 7.844009e-02 FALSE  
## 35 1.0000000 FALSE 1.41564260 7.844009e-02 FALSE  
## 36 1.0000000 FALSE 1.41564260 7.844009e-02 FALSE  
## 37 1.0000000 FALSE 1.41564260 7.844009e-02 FALSE  
## 38 1.0000000 FALSE 1.36792226 8.566821e-02 FALSE  
## 39 1.0000000 FALSE 1.29035100 9.846441e-02 FALSE  
## 40 1.0000000 FALSE 1.29035100 9.846441e-02 FALSE  
## 41 1.0000000 FALSE 1.29035100 9.846441e-02 FALSE  
## 42 1.0000000 FALSE 1.29035100 9.846441e-02 FALSE  
## 43 1.0000000 FALSE 1.29035100 9.846441e-02 FALSE  
## 44 1.0000000 FALSE 1.29035100 9.846441e-02 FALSE  
## 45 1.0000000 FALSE 1.27521031 1.011174e-01 FALSE  
## 46 1.0000000 FALSE 1.24846382 1.059306e-01 FALSE  
## 47 1.0000000 FALSE 1.22524017 1.102424e-01 FALSE  
## 48 1.0000000 FALSE 1.18041952 1.189167e-01 FALSE  
## 49 1.0000000 FALSE 1.15586734 1.238677e-01 FALSE  
## 50 1.0000000 FALSE 1.15586734 1.238677e-01 FALSE  
## 51 1.0000000 FALSE 1.15586734 1.238677e-01 FALSE  
## 52 1.0000000 FALSE 1.15586734 1.238677e-01 FALSE  
## 53 1.0000000 FALSE 1.15586734 1.238677e-01 FALSE  
## 54 1.0000000 FALSE 1.15586734 1.238677e-01 FALSE  
## 55 1.0000000 FALSE 1.15586734 1.238677e-01 FALSE  
## 56 1.0000000 FALSE 1.15586734 1.238677e-01 FALSE  
## 57 1.0000000 FALSE 1.15586734 1.238677e-01 FALSE  
## 58 1.0000000 FALSE 1.15586734 1.238677e-01 FALSE  
## 59 1.0000000 FALSE 1.15586734 1.238677e-01 FALSE  
## 60 1.0000000 FALSE 1.15586734 1.238677e-01 FALSE  
## 61 1.0000000 FALSE 1.15586734 1.238677e-01 FALSE  
## 62 1.0000000 FALSE 1.15586734 1.238677e-01 FALSE  
## 63 1.0000000 FALSE 1.15586734 1.238677e-01 FALSE  
## 64 1.0000000 FALSE 1.15586734 1.238677e-01 FALSE  
## 65 1.0000000 FALSE 1.15586734 1.238677e-01 FALSE  
## 66 1.0000000 FALSE 1.15586734 1.238677e-01 FALSE  
## 67 1.0000000 FALSE 1.15586734 1.238677e-01 FALSE  
## 68 1.0000000 FALSE 1.15586734 1.238677e-01 FALSE  
## 69 1.0000000 FALSE 1.15586734 1.238677e-01 FALSE  
## 70 1.0000000 FALSE 1.15586734 1.238677e-01 FALSE  
## 71 1.0000000 FALSE 1.15586734 1.238677e-01 FALSE  
## 72 1.0000000 FALSE 1.15586734 1.238677e-01 FALSE  
## 73 1.0000000 FALSE 1.14134285 1.268636e-01 FALSE  
## 74 1.0000000 FALSE 1.11200508 1.330680e-01 FALSE  
## 75 1.0000000 FALSE 1.07417393 1.413724e-01 FALSE  
## 76 1.0000000 FALSE 1.06845742 1.426571e-01 FALSE  
## 77 1.0000000 FALSE 1.06803308 1.427528e-01 FALSE  
## 78 1.0000000 FALSE 1.05356718 1.460406e-01 FALSE  
## 79 1.0000000 FALSE 1.05356718 1.460406e-01 FALSE  
## 80 1.0000000 FALSE 1.05356718 1.460406e-01 FALSE  
## 81 1.0000000 FALSE 1.05356718 1.460406e-01 FALSE  
## 82 1.0000000 FALSE 1.05356718 1.460406e-01 FALSE  
## 83 1.0000000 FALSE 1.05356718 1.460406e-01 FALSE  
## 84 1.0000000 FALSE 1.05356718 1.460406e-01 FALSE  
## 85 1.0000000 FALSE 1.05356718 1.460406e-01 FALSE  
## 86 1.0000000 FALSE 1.05356718 1.460406e-01 FALSE  
## 87 1.0000000 FALSE 1.05356718 1.460406e-01 FALSE  
## 88 1.0000000 FALSE 1.05356718 1.460406e-01 FALSE  
## 89 1.0000000 FALSE 1.05356718 1.460406e-01 FALSE  
## 90 1.0000000 FALSE 1.05356718 1.460406e-01 FALSE  
## 91 1.0000000 FALSE 1.05356718 1.460406e-01 FALSE  
## 92 1.0000000 FALSE 1.05356718 1.460406e-01 FALSE  
## 93 1.0000000 FALSE 1.05356718 1.460406e-01 FALSE  
## 94 1.0000000 FALSE 1.05356718 1.460406e-01 FALSE  
## 95 1.0000000 FALSE 1.05356718 1.460406e-01 FALSE  
## 96 1.0000000 FALSE 1.05356718 1.460406e-01 FALSE  
## 97 1.0000000 FALSE 1.05356718 1.460406e-01 FALSE  
## 98 1.0000000 FALSE 1.05356718 1.460406e-01 FALSE  
## 99 1.0000000 FALSE 1.05356718 1.460406e-01 FALSE  
## 100 1.0000000 FALSE 1.05356718 1.460406e-01 FALSE  
## 101 1.0000000 FALSE 1.05356718 1.460406e-01 FALSE  
## 102 1.0000000 FALSE 1.05356718 1.460406e-01 FALSE  
## 103 1.0000000 FALSE 1.05356718 1.460406e-01 FALSE  
## 104 1.0000000 FALSE 1.05356718 1.460406e-01 FALSE  
## 105 1.0000000 FALSE 1.05356718 1.460406e-01 FALSE  
## 106 1.0000000 FALSE 1.05356718 1.460406e-01 FALSE  
## 107 1.0000000 FALSE 1.05356718 1.460406e-01 FALSE  
## 108 1.0000000 FALSE 1.05356718 1.460406e-01 FALSE  
## 109 1.0000000 FALSE 1.05356718 1.460406e-01 FALSE  
## 110 1.0000000 FALSE 1.05356718 1.460406e-01 FALSE  
## 111 1.0000000 FALSE 1.05356718 1.460406e-01 FALSE  
## 112 1.0000000 FALSE 1.05356718 1.460406e-01 FALSE  
## 113 1.0000000 FALSE 1.05356718 1.460406e-01 FALSE  
## 114 1.0000000 FALSE 1.05356718 1.460406e-01 FALSE  
## 115 1.0000000 FALSE 1.05356718 1.460406e-01 FALSE  
## 116 1.0000000 FALSE 1.05356718 1.460406e-01 FALSE  
## 117 1.0000000 FALSE 1.05356718 1.460406e-01 FALSE  
## 118 1.0000000 FALSE 1.05356718 1.460406e-01 FALSE  
## 119 1.0000000 FALSE 1.05356718 1.460406e-01 FALSE  
## 120 1.0000000 FALSE 1.05356718 1.460406e-01 FALSE  
## 121 1.0000000 FALSE 1.05356718 1.460406e-01 FALSE  
## 122 1.0000000 FALSE 1.05356718 1.460406e-01 FALSE  
## 123 1.0000000 FALSE 1.05356718 1.460406e-01 FALSE  
## 124 1.0000000 FALSE 1.03689675 1.498920e-01 FALSE  
## 125 1.0000000 FALSE 1.03689675 1.498920e-01 FALSE  
## 126 1.0000000 FALSE 1.02222902 1.533363e-01 FALSE  
## 127 1.0000000 FALSE 1.00559412 1.573054e-01 FALSE  
## 128 1.0000000 FALSE 0.92122269 1.784671e-01 FALSE  
## 129 1.0000000 FALSE 0.91837805 1.792105e-01 FALSE  
## 130 1.0000000 FALSE 0.91501968 1.800906e-01 FALSE  
## 131 1.0000000 FALSE 0.85962417 1.949981e-01 FALSE  
## 132 1.0000000 FALSE 0.82529916 2.046009e-01 FALSE  
## 133 1.0000000 FALSE 0.81093961 2.087002e-01 FALSE  
## 134 1.0000000 FALSE 0.79533872 2.132082e-01 FALSE  
## 135 1.0000000 FALSE 0.78954131 2.148978e-01 FALSE  
## 136 1.0000000 FALSE 0.78630633 2.158440e-01 FALSE  
## 137 1.0000000 FALSE 0.78630633 2.158440e-01 FALSE  
## 138 1.0000000 FALSE 0.74957248 2.267561e-01 FALSE  
## 139 1.0000000 FALSE 0.74957248 2.267561e-01 FALSE  
## 140 1.0000000 FALSE 0.74394590 2.284546e-01 FALSE  
## 141 1.0000000 FALSE 0.73247505 2.319393e-01 FALSE  
## 142 1.0000000 FALSE 0.70122094 2.415826e-01 FALSE  
## 143 1.0000000 FALSE 0.68014478 2.482064e-01 FALSE  
## 144 1.0000000 FALSE 0.67530234 2.497418e-01 FALSE  
## 145 1.0000000 FALSE 0.64939134 2.580427e-01 FALSE  
## 146 1.0000000 FALSE 0.64939134 2.580427e-01 FALSE  
## 147 1.0000000 FALSE 0.64939134 2.580427e-01 FALSE  
## 148 1.0000000 FALSE 0.60727503 2.718342e-01 FALSE  
## 149 1.0000000 FALSE 0.58019231 2.808925e-01 FALSE  
## 150 1.0000000 FALSE 0.56758725 2.851576e-01 FALSE  
## 151 1.0000000 FALSE 0.52604919 2.994270e-01 FALSE  
## 152 1.0000000 FALSE 0.52604919 2.994270e-01 FALSE  
## 153 1.0000000 FALSE 0.52591561 2.994734e-01 FALSE  
## 154 1.0000000 FALSE 0.50622589 3.063490e-01 FALSE  
## 155 1.0000000 FALSE 0.49583808 3.100043e-01 FALSE  
## 156 1.0000000 FALSE 0.49583808 3.100043e-01 FALSE  
## 157 1.0000000 FALSE 0.47895736 3.159845e-01 FALSE  
## 158 1.0000000 FALSE 0.43776906 3.307769e-01 FALSE  
## 159 1.0000000 FALSE 0.42940829 3.338131e-01 FALSE  
## 160 1.0000000 FALSE 0.42940829 3.338131e-01 FALSE  
## 161 1.0000000 FALSE 0.42940829 3.338131e-01 FALSE  
## 162 1.0000000 FALSE 0.41720229 3.382652e-01 FALSE  
## 163 1.0000000 FALSE 0.40528637 3.426335e-01 FALSE  
## 164 1.0000000 FALSE 0.37820538 3.526390e-01 FALSE  
## 165 1.0000000 FALSE 0.37820538 3.526390e-01 FALSE  
## 166 1.0000000 FALSE 0.37820538 3.526390e-01 FALSE  
## 167 1.0000000 FALSE 0.37820538 3.526390e-01 FALSE  
## 168 1.0000000 FALSE 0.37820538 3.526390e-01 FALSE  
## 169 1.0000000 FALSE 0.37820538 3.526390e-01 FALSE  
## 170 1.0000000 FALSE 0.37562915 3.535963e-01 FALSE  
## 171 1.0000000 FALSE 0.37534887 3.537005e-01 FALSE  
## 172 1.0000000 FALSE 0.35491001 3.613285e-01 FALSE  
## 173 1.0000000 FALSE 0.33867399 3.674277e-01 FALSE  
## 174 1.0000000 FALSE 0.33867399 3.674277e-01 FALSE  
## 175 1.0000000 FALSE 0.32036470 3.743459e-01 FALSE  
## 176 1.0000000 FALSE 0.30908277 3.786293e-01 FALSE  
## 177 1.0000000 FALSE 0.30599887 3.798028e-01 FALSE  
## 178 1.0000000 FALSE 0.29140603 3.853704e-01 FALSE  
## 179 1.0000000 FALSE 0.27258538 3.925860e-01 FALSE  
## 180 1.0000000 FALSE 0.27258538 3.925860e-01 FALSE  
## 181 1.0000000 FALSE 0.26560992 3.952698e-01 FALSE  
## 182 1.0000000 FALSE 0.26560992 3.952698e-01 FALSE  
## 183 1.0000000 FALSE 0.23037669 4.088995e-01 FALSE  
## 184 1.0000000 FALSE 0.18321924 4.273130e-01 FALSE  
## 185 1.0000000 FALSE 0.17844902 4.291852e-01 FALSE  
## 186 1.0000000 FALSE 0.17003624 4.324908e-01 FALSE  
## 187 1.0000000 FALSE 0.17003624 4.324908e-01 FALSE  
## 188 1.0000000 FALSE 0.17003624 4.324908e-01 FALSE  
## 189 1.0000000 FALSE 0.17003624 4.324908e-01 FALSE  
## 190 1.0000000 FALSE 0.17003624 4.324908e-01 FALSE  
## 191 1.0000000 FALSE 0.17003624 4.324908e-01 FALSE  
## 192 1.0000000 FALSE 0.14570301 4.420779e-01 FALSE  
## 193 1.0000000 FALSE 0.14570301 4.420779e-01 FALSE  
## 194 1.0000000 FALSE 0.12535064 4.501230e-01 FALSE  
## 195 1.0000000 FALSE 0.10681744 4.574669e-01 FALSE  
## 196 1.0000000 FALSE 0.10411068 4.585408e-01 FALSE  
## 197 1.0000000 FALSE 0.10070845 4.598910e-01 FALSE  
## 198 1.0000000 FALSE 0.10008537 4.601383e-01 FALSE  
## 199 1.0000000 FALSE 0.09420412 4.624735e-01 FALSE  
## 200 1.0000000 FALSE 0.09420412 4.624735e-01 FALSE  
## 201 1.0000000 FALSE 0.08863629 4.646855e-01 FALSE  
## 202 1.0000000 FALSE 0.08863629 4.646855e-01 FALSE  
## 203 1.0000000 FALSE 0.07961702 4.682709e-01 FALSE  
## 204 1.0000000 FALSE 0.07222449 4.712116e-01 FALSE  
## 205 1.0000000 FALSE 0.06167108 4.754124e-01 FALSE  
## 206 1.0000000 FALSE 0.05296938 4.788782e-01 FALSE  
## 207 1.0000000 FALSE 0.05035422 4.799201e-01 FALSE  
## 208 1.0000000 FALSE 0.05035422 4.799201e-01 FALSE  
##   
##   
## Summary statistics:  
##   
## Total Chi squared = 990.2424   
## Total degrees of freedom = 466   
## p = 0   
## Sum of counts = 2138   
##   
## Levels:  
##   
## control variety x   
## 2 2 156

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

## [1] "Australia jealous" "Australia threatened" "Australia paranoid"   
## [4] "Australia other" "Ireland excuses" "Ireland coincidence"   
## [7] "Australia hypocrite" "Australia generalise" "Australia entitled"   
## [10] "Australia creepy" "Ireland agenda" "Australia agenda"

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).

## CIT

Process data

citdata <- xmuch %>%  
 dplyr::filter(control == "test") %>%  
 dplyr::select(variety, pos, element, emo, polarity, x) %>%  
 dplyr::rename(X = x,  
 Variety = variety,  
 Polarity = polarity,  
 Emotionality = emo,  
 Unit = element,  
 POS = pos) %>%  
 dplyr::group\_by(Variety, POS, Unit, Emotionality, Polarity, X) %>%  
 dplyr::summarise(Frequency = n()) %>%  
 dplyr::ungroup()

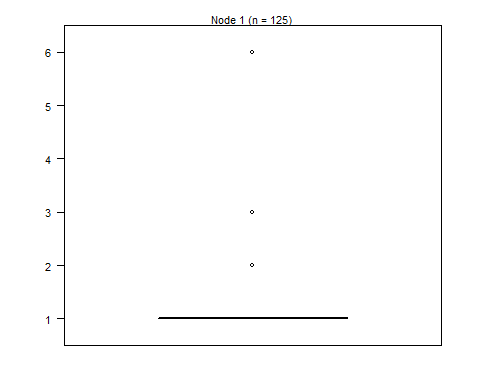
## `summarise()` has grouped output by 'Variety', 'POS', 'Unit', 'Emotionality', 'Polarity'. You can override using the `.groups` argument.

# inspect  
head(citdata)

## # A tibble: 6 x 7  
## Variety POS Unit Emotionality Polarity X Frequency  
## <chr> <chr> <chr> <chr> <chr> <chr> <int>  
## 1 Australia Adjective phrase nonemotional nonemotional 14 tubesteak m~ 1  
## 2 Australia Adjective phrase nonemotional nonemotional arrogant and a~ 1  
## 3 Australia Adjective phrase nonemotional nonemotional hey peter bitt~ 1  
## 4 Australia Adjective phrase nonemotional nonemotional wow desperate 1  
## 5 Australia Adjective word emotional negative condescending 1  
## 6 Australia Adjective word emotional negative crazy 1

Perform CIT analysis with the *party* (Hothorn, Hornik, and Zeileis [2006](#ref-party)) and the *partykit* (Hothorn and Zeileis [2015](#ref-partykit)) packages

# set.seed  
set.seed(111)   
# apply bonferroni correction (1 minus alpha multiplied by n of predictors)  
control = ctree\_control(mincriterion = 1-(.05\*ncol(citdata)-1))  
# convert character strings to factors  
citdata <- citdata %>%  
 dplyr::mutate\_if(is.character, factor)  
# create initial conditional inference tree model  
citd.ctree <- partykit::ctree(Frequency ~ Variety + POS + Unit + Emotionality + Polarity,  
 data = citdata)  
plot(citd.ctree, gp = gpar(fontsize = 8)) # plot final ctree



The CIT analysis does nto confirm significant stratification in the data.

# 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] stats4 grid stats graphics grDevices utils datasets   
## [8] methods base   
##   
## other attached packages:  
## [1] effectsize\_0.5 partykit\_1.2-15 libcoin\_1.0-9 party\_1.3-9   
## [5] strucchange\_1.5-2 sandwich\_3.0-1 zoo\_1.8-9 modeltools\_0.2-23  
## [9] mvtnorm\_1.1-3 vcd\_1.4-9 cfa\_0.10-0 tidytext\_0.3.2   
## [13] quanteda\_3.2.0 flextable\_0.6.10 readxl\_1.3.1 here\_1.0.1   
## [17] forcats\_0.5.1 stringr\_1.4.0 dplyr\_1.0.7 purrr\_0.3.4   
## [21] readr\_2.1.1 tidyr\_1.1.4 tibble\_3.1.6 ggplot2\_3.3.5   
## [25] tidyverse\_1.3.1   
##   
## loaded via a namespace (and not attached):  
## [1] TH.data\_1.1-0 colorspace\_2.0-2 ellipsis\_0.3.2 rprojroot\_2.0.2   
## [5] estimability\_1.3 parameters\_0.15.0 base64enc\_0.1-3 fs\_1.5.2   
## [9] rstudioapi\_0.13 SnowballC\_0.7.0 fansi\_0.5.0 lubridate\_1.8.0   
## [13] coin\_1.4-2 xml2\_1.3.3 codetools\_0.2-18 splines\_4.1.1   
## [17] knitr\_1.37 Formula\_1.2-4 jsonlite\_1.7.2 broom\_0.7.10   
## [21] dbplyr\_2.1.1 compiler\_4.1.1 httr\_1.4.2 emmeans\_1.7.1-1   
## [25] backports\_1.4.1 assertthat\_0.2.1 Matrix\_1.4-0 fastmap\_1.1.0   
## [29] cli\_3.1.0 htmltools\_0.5.2 tools\_4.1.1 coda\_0.19-4   
## [33] gtable\_0.3.0 glue\_1.6.0 fastmatch\_1.1-3 Rcpp\_1.0.7   
## [37] cellranger\_1.1.0 vctrs\_0.3.8 insight\_0.14.5 lmtest\_0.9-39   
## [41] inum\_1.0-4 xfun\_0.29 stopwords\_2.3 rvest\_1.0.2   
## [45] lifecycle\_1.0.1 MASS\_7.3-54 scales\_1.1.1 hms\_1.1.1   
## [49] parallel\_4.1.1 RColorBrewer\_1.1-2 yaml\_2.2.1 gdtools\_0.2.3   
## [53] rpart\_4.1-15 stringi\_1.7.6 highr\_0.9 bayestestR\_0.11.5   
## [57] tokenizers\_0.2.1 zip\_2.2.0 rlang\_0.4.12 pkgconfig\_2.0.3   
## [61] systemfonts\_1.0.3 matrixStats\_0.61.0 evaluate\_0.14 lattice\_0.20-45   
## [65] tidyselect\_1.1.1 magrittr\_2.0.1 R6\_2.5.1 generics\_0.1.1   
## [69] multcomp\_1.4-17 DBI\_1.1.2 pillar\_1.6.4 haven\_2.4.3   
## [73] withr\_2.4.3 datawizard\_0.2.1 survival\_3.2-13 performance\_0.8.0   
## [77] janeaustenr\_0.1.5 modelr\_0.1.8 crayon\_1.4.2 uuid\_1.0-3   
## [81] utf8\_1.2.2 tzdb\_0.2.0 rmarkdown\_2.11 officer\_0.4.1   
## [85] data.table\_1.14.2 reprex\_2.0.1.9000 digest\_0.6.28 xtable\_1.8-4   
## [89] RcppParallel\_5.1.4 munsell\_0.5.0

# 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.

Hothorn, Torsten, Kurt Hornik, and Achim Zeileis. 2006. “Unbiased Recursive Partitioning: A Conditional Inference Framework.” *Journal of Computational and Graphical Statistics* 15 (3): 651–74.

Hothorn, Torsten, and Achim Zeileis. 2015. “Partykit: A Modular Toolkit for Recursive Partytioning in R.” *Journal of Machine Learning Research* 16: 3905–9.

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>.