Contact helps dispreferred combinations of typological features to survive: geospatial evidence

Statistical analyses

Preliminaries

```r
# required packages
require(tidyverse)
```

```
## Loading required package: tidyverse
```

```
## -- Attaching core tidyverse packages ------------------------------------------------
## v dplyr     1.1.4     v readr     2.1.5
## v forcats   1.0.0     v stringr   1.5.1
## v ggplot2   3.5.1     v tibble    3.2.1
## v lubridate 1.9.3     v tidyr     1.3.1
## v purrr     1.0.2
## -- Conflicts ------------------------------------------------------------------------
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors
```

```r
require(broom)
```

```
## Loading required package: broom
```

```r
require(pixiedust)
```

```
## Loading required package: pixiedust
```

```r
require(emmeans)
```

```
## Loading required package: emmeans
## Welcome to emmeans.
## Caution: You lose important information if you filter this package's results.
## See '? untidy'
```

```r
# create results/tables, if it doesn't exist already
try(dir.create("../results/tables", recursive=TRUE))
```

```
## Warning in dir.create("../results/tables", recursive = TRUE):
## '../results/tables' already exists
```

```r
# load (and rename) dataframe "combined", which contains all our data
load("../results/combined.RData")
data <- combined
data <- data[data$pair != "PolQ & NegM", ]
data <- data[data$pair != "Gen & Pas", ]

# make "non-interacting" the reference level of "status" factor
```

```
data$status <- relevel(data$status, ref="non-interacting")

# inflection points
infl <- read.csv("../results/tables/inflection_points.csv")
ip_wals <- round(mean(infl[infl$dataset == "WALS" & !is.na(infl$inflpoint), ]$inflpoint))
ip_grambank <- round(mean(infl[infl$dataset == "Grambank" & !is.na(infl$inflpoint), ]$inflpoint))

print(ip_wals)
```

```
## [1] 12
```

```
print(ip_grambank)
```

```
## [1] 14
```

```
# restrict to final choice of k
wals <- data[data$dataset == "WALS" & data$k == ip_wals, ]
gram <- data[data$dataset == "Grambank" & data$k == ip_grambank, ]
```

Basic model: comparison of Δ between typologies of different statuses

Under-represented types (Δ^-)

```
mod_w <- lm(Delta_under ~ status+abs(phi), data=wals)
mod_g <- lm(Delta_under ~ status+abs(phi), data=gram)

#mod_w %>% dust %>% sprinkle(round=5) %>% write.csv(file="../results/tables/mod1_under_wals.csv", row.n
#mod_g %>% dust %>% sprinkle(round=5) %>% write.csv(file="../results/tables/mod1_under_grambank.csv", r

print(summary(mod_w))
```

```
##
## Call:
## lm(formula = Delta_under ~ status + abs(phi), data = wals)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.11904 -0.05447 -0.01653  0.04151  0.17389
##
## Coefficients:
##                   Estimate Std. Error t value Pr(>|t|)
## (Intercept)        0.07189    0.02524   2.848  0.00691 **
## statusinteracting  0.08678    0.04122   2.105  0.04158 *
## statusunknown      0.04011    0.02760   1.453  0.15400
## abs(phi)          -0.03246    0.07360  -0.441  0.66153
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.07878 on 40 degrees of freedom
## Multiple R-squared:  0.1338, Adjusted R-squared:  0.06879
## F-statistic: 2.059 on 3 and 40 DF,  p-value: 0.121
```

```
print(summary(mod_g))
```

```
##
## Call:
## lm(formula = Delta_under ~ status + abs(phi), data = gram)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.20718 -0.05024 -0.01081  0.06143  0.16423
##
## Coefficients:
##                   Estimate Std. Error t value Pr(>|t|)
## (Intercept)        0.11137    0.02228   4.999 1.19e-05 ***
## statusinteracting  0.13096    0.05350   2.448   0.0188 *
## statusunknown      0.06099    0.03165   1.927   0.0611 .
## abs(phi)          -0.06721    0.07100  -0.947   0.3495
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.08805 on 40 degrees of freedom
## Multiple R-squared:  0.1484, Adjusted R-squared:  0.08455
## F-statistic: 2.324 on 3 and 40 DF,  p-value: 0.08946
```

Over-represented types (Δ^+)

```
mod_w <- lm(Delta_over ~ status+abs(phi), data=wals)
mod_g <- lm(Delta_over ~ status+abs(phi), data=gram)

#mod_w %>% dust %>% sprinkle(round=5) %>% write.csv(file="../results/tables/mod1_over_wals.csv", row.na
#mod_g %>% dust %>% sprinkle(round=5) %>% write.csv(file="../results/tables/mod1_over_grambank.csv", ro

print(summary(mod_w))
```

```
##
## Call:
## lm(formula = Delta_over ~ status + abs(phi), data = wals)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.12155 -0.01796  0.01736  0.02645  0.05006
##
## Coefficients:
##                    Estimate Std. Error t value Pr(>|t|)
## (Intercept)       -0.021804   0.014417  -1.512    0.138
## statusinteracting -0.004750   0.023542  -0.202    0.841
## statusunknown     -0.026796   0.015765  -1.700    0.097 .
## abs(phi)          -0.009045   0.042038  -0.215    0.831
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.045 on 40 degrees of freedom
## Multiple R-squared:  0.07388,    Adjusted R-squared:  0.00442
## F-statistic: 1.064 on 3 and 40 DF,  p-value: 0.3753
```

3

```
print(summary(mod_g))
```

```
##
## Call:
## lm(formula = Delta_over ~ status + abs(phi), data = gram)
##
## Residuals:
##       Min        1Q    Median        3Q       Max
## -0.138736 -0.017352  0.006922  0.024847  0.050873
##
## Coefficients:
##                    Estimate Std. Error t value Pr(>|t|)
## (Intercept)        -0.02612    0.01026  -2.546   0.0149 *
## statusinteracting  -0.01980    0.02463  -0.804   0.4263
## statusunknown      -0.01705    0.01457  -1.170   0.2490
## abs(phi)           -0.01534    0.03269  -0.469   0.6414
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.04054 on 40 degrees of freedom
## Multiple R-squared:  0.0709, Adjusted R-squared:  0.001214
## F-statistic: 1.017 on 3 and 40 DF,  p-value: 0.3951
```

Model 2: model comparison between φ and φ_c as predictors

Under-represented types

```
mod_w <- lm(Delta_under ~ abs(phi), data=wals)
mod_wc <- lm(Delta_under ~ abs(corrected_phi), data=wals)

print(summary(mod_w))
```

```
##
## Call:
## lm(formula = Delta_under ~ abs(phi), data = wals)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.11598 -0.06214 -0.02978  0.04249  0.17977
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.07867    0.02165   3.633 0.000756 ***
## abs(phi)     0.06169    0.05523   1.117 0.270340
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.0814 on 42 degrees of freedom
## Multiple R-squared:  0.02885,    Adjusted R-squared:  0.005728
## F-statistic: 1.248 on 1 and 42 DF,  p-value: 0.2703
```

```
print(summary(mod_wc))
```

```
##
```

```
## Call:
## lm(formula = Delta_under ~ abs(corrected_phi), data = wals)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.10239 -0.06570 -0.01390  0.03306  0.19530
##
## Coefficients:
##                     Estimate Std. Error t value Pr(>|t|)
## (Intercept)          0.05779    0.01722   3.357  0.00168 **
## abs(corrected_phi)   0.21122    0.06761   3.124  0.00323 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.07441 on 42 degrees of freedom
## Multiple R-squared:  0.1886, Adjusted R-squared:  0.1693
## F-statistic: 9.761 on 1 and 42 DF,  p-value: 0.003227
```

```
print(AIC(mod_w))
```

```
## [1] -91.91508
```

```
print(AIC(mod_wc))
```

```
## [1] -99.82139
```

```
mod_g <- lm(Delta_under ~ abs(phi), data=gram)
mod_gc <- lm(Delta_under ~ abs(corrected_phi), data=gram)

print(summary(mod_g))
```

```
##
## Call:
## lm(formula = Delta_under ~ abs(phi), data = gram)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.17307 -0.05184 -0.00917  0.05921  0.19476
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.13824    0.01838   7.523 2.62e-09 ***
## abs(phi)     0.03576    0.06173   0.579    0.566
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.09275 on 42 degrees of freedom
## Multiple R-squared:  0.007925,  Adjusted R-squared:  -0.0157
## F-statistic: 0.3355 on 1 and 42 DF,  p-value: 0.5655
```

```
print(summary(mod_gc))
```

```
##
## Call:
## lm(formula = Delta_under ~ abs(corrected_phi), data = gram)
##
## Residuals:
```

```
##       Min       1Q   Median       3Q      Max
## -0.18244 -0.04955 -0.02279  0.04277  0.20238
##
## Coefficients:
##                   Estimate Std. Error t value Pr(>|t|)
## (Intercept)        0.12243    0.01868   6.554 6.35e-08 ***
## abs(corrected_phi) 0.15744    0.08915   1.766   0.0847 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.08984 on 42 degrees of freedom
## Multiple R-squared:  0.06912,    Adjusted R-squared:  0.04695
## F-statistic: 3.119 on 1 and 42 DF,  p-value: 0.08468
```

```
print(AIC(mod_g))
```

```
## [1] -80.4348
```

```
print(AIC(mod_gc))
```

```
## [1] -83.23613
```

```
#mod_w %>% dust %>% sprinkle(round=5) %>% write.csv(file="../results/tables/mod2_under_wals.csv", row.n
#mod_g %>% dust %>% sprinkle(round=5) %>% write.csv(file="../results/tables/mod2_under_grambank.csv", r
#mod_wc %>% dust %>% sprinkle(round=5) %>% write.csv(file="../results/tables/mod2_under_corrected_wals.
#mod_gc %>% dust %>% sprinkle(round=5) %>% write.csv(file="../results/tables/mod2_under_corrected_gramb
```

Over-represented types

```
mod_w <- lm(Delta_over ~ abs(phi), data=wals)
mod_wc <- lm(Delta_over ~ abs(corrected_phi), data=wals)

print(summary(mod_w))
```

```
##
## Call:
## lm(formula = Delta_over ~ abs(phi), data = wals)
##
## Residuals:
##       Min       1Q   Median       3Q      Max
## -0.13799 -0.03099  0.01586  0.03483  0.03737
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -0.035905   0.012137  -2.958  0.00506 **
## abs(phi)    -0.001117   0.030959  -0.036  0.97140
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.04563 on 42 degrees of freedom
## Multiple R-squared:  3.097e-05,  Adjusted R-squared:  -0.02378
## F-statistic: 0.001301 on 1 and 42 DF,  p-value: 0.9714
```

```
print(summary(mod_wc))
```

```
##
```

```
## Call:
## lm(formula = Delta_over ~ abs(corrected_phi), data = wals)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.13856 -0.03122  0.01634  0.03499  0.03937
##
## Coefficients:
##                    Estimate Std. Error t value Pr(>|t|)
## (Intercept)        -0.03504    0.01056  -3.320  0.00187 **
## abs(corrected_phi) -0.00636    0.04145  -0.153  0.87879
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.04562 on 42 degrees of freedom
## Multiple R-squared:  0.0005602,  Adjusted R-squared:  -0.02324
## F-statistic: 0.02354 on 1 and 42 DF,  p-value: 0.8788
```

```
print(AIC(mod_w))
```

```
## [1] -142.8538
```

```
print(AIC(mod_wc))
```

```
## [1] -142.8771
```

```
mod_g <- lm(Delta_over ~ abs(phi), data=gram)
mod_gc <- lm(Delta_over ~ abs(corrected_phi), data=gram)

print(summary(mod_g))
```

```
##
## Call:
## lm(formula = Delta_over ~ abs(phi), data = gram)
##
## Residuals:
##       Min        1Q    Median        3Q       Max
## -0.140011 -0.018107  0.006116  0.031230  0.051168
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) -0.033590   0.007978  -4.210 0.000132 ***
## abs(phi)    -0.034434   0.026798  -1.285 0.205853
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.04026 on 42 degrees of freedom
## Multiple R-squared:  0.03782,    Adjusted R-squared:  0.01492
## F-statistic: 1.651 on 1 and 42 DF,  p-value: 0.2059
```

```
print(summary(mod_gc))
```

```
##
## Call:
## lm(formula = Delta_over ~ abs(corrected_phi), data = gram)
##
## Residuals:
```

```
##       Min       1Q   Median       3Q      Max
## -0.14915 -0.01464  0.00091  0.03577  0.04129
##
## Coefficients:
##                     Estimate Std. Error t value Pr(>|t|)
## (Intercept)        -0.039359   0.008533  -4.613 3.71e-05 ***
## abs(corrected_phi) -0.006116   0.040722  -0.150    0.881
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.04104 on 42 degrees of freedom
## Multiple R-squared:  0.0005368,  Adjusted R-squared:  -0.02326
## F-statistic: 0.02256 on 1 and 42 DF,  p-value: 0.8813
```

```
print(AIC(mod_g))
```

```
## [1] -153.864
```

```
print(AIC(mod_gc))
```

```
## [1] -152.191
```