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

Statistical analyses

Preliminaries

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

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

```
## -- Attaching packages --------------------------------------------------------- tidyverse 1.3.2 --
## v ggplot2 3.4.1      v purrr   1.0.1
## v tibble  3.1.8      v dplyr   1.0.10
## v tidyr   1.3.0      v stringr 1.5.0
## v readr   2.1.4      v forcats 1.0.0
## -- Conflicts ------------------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```r
require(broom)
```

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

```r
require(pixiedust)
```

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

```r
require(emmeans)
```

```
## Loading required package: emmeans
```

```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))
```

```r
ip_grambank <- round(mean(infl[infl$dataset == "Grambank" & !is.na(infl$inflpoint), ]$inflpoint))

print(ip_wals)
```

```
## [1] 22
```

```r
print(ip_grambank)
```

```
## [1] 28
```

```r
# 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^-)

```r
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
## -0.07606 -0.03707 -0.01902  0.03307  0.13410
##
## Coefficients:
##                   Estimate Std. Error t value Pr(>|t|)
## (Intercept)        0.04648    0.01756   2.647   0.0115 *
## statusinteracting  0.06206    0.02867   2.165   0.0364 *
## statusunknown      0.02837    0.01920   1.478   0.1474
## abs(phi)          -0.04996    0.05119  -0.976   0.3350
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.05479 on 40 degrees of freedom
## Multiple R-squared:  0.1174, Adjusted R-squared:  0.05122
## F-statistic: 1.774 on 3 and 40 DF,  p-value: 0.1677
```

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

```
##
## Call:
## lm(formula = Delta_under ~ status + abs(phi), data = gram)
##
```

```
## Residuals:
##       Min        1Q    Median        3Q       Max
## -0.109673 -0.046569  0.002067  0.032171  0.123436
##
## Coefficients:
##                    Estimate Std. Error t value Pr(>|t|)
## (Intercept)         0.05405    0.01560   3.464  0.00128 **
## statusinteracting   0.11019    0.03747   2.941  0.00542 **
## statusunknown       0.05910    0.02217   2.666  0.01102 *
## abs(phi)           -0.06055    0.04972  -1.218  0.23046
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.06167 on 40 degrees of freedom
## Multiple R-squared:  0.2153, Adjusted R-squared:  0.1565
## F-statistic: 3.659 on 3 and 40 DF,  p-value: 0.02018
```

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
## -0.080568 -0.001464  0.009259  0.012329  0.028303
##
## Coefficients:
##                    Estimate Std. Error t value Pr(>|t|)
## (Intercept)        -0.009582   0.008332  -1.150   0.2569
## statusinteracting  -0.002368   0.013604  -0.174   0.8627
## statusunknown      -0.015562   0.009110  -1.708   0.0954 .
## abs(phi)           -0.007199   0.024293  -0.296   0.7685
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.026 on 40 degrees of freedom
## Multiple R-squared:  0.0756, Adjusted R-squared:  0.006265
## F-statistic:  1.09 on 3 and 40 DF,  p-value: 0.3643
```

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

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

```
##        Min        1Q     Median        3Q        Max
## -0.049143 -0.012222   0.007617   0.009568   0.022617
##
## Coefficients:
##                      Estimate Std. Error t value Pr(>|t|)
## (Intercept)         -0.008292   0.004829  -1.717   0.0937 .
## statusinteracting   -0.008109   0.011597  -0.699   0.4884
## statusunknown       -0.007814   0.006861  -1.139   0.2615
## abs(phi)            -0.012186   0.015390  -0.792   0.4332
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.01909 on 40 degrees of freedom
## Multiple R-squared:  0.0886, Adjusted R-squared:  0.02025
## F-statistic: 1.296 on 3 and 40 DF,  p-value: 0.289
```

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
## -0.06625 -0.04280 -0.02075   0.02071   0.13811
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.05115    0.01510   3.387  0.00155 **
## abs(phi)     0.01755    0.03852   0.456  0.65104
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.05678 on 42 degrees of freedom
## Multiple R-squared:  0.004917,   Adjusted R-squared:  -0.01878
## F-statistic: 0.2076 on 1 and 42 DF,  p-value: 0.651
```

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

```
##
## Call:
## lm(formula = Delta_under ~ abs(corrected_phi), data = wals)
##
## Residuals:
##       Min        1Q    Median        3Q       Max
## -0.07028 -0.04011 -0.01530   0.02867   0.13923
##
```

```
## Coefficients:
##                  Estimate Std. Error t value Pr(>|t|)
## (Intercept)       0.03233    0.01219   2.652   0.0112 *
## abs(corrected_phi) 0.12676    0.04787   2.648   0.0114 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.05269 on 42 degrees of freedom
## Multiple R-squared:  0.143,  Adjusted R-squared:  0.1226
## F-statistic:  7.01 on 1 and 42 DF,  p-value: 0.01136
```

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

```
## [1] -123.6155
```

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

```
## [1] -130.1904
```

```
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
## -0.09368 -0.05300 -0.01819  0.04627  0.14279
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.08005    0.01339   5.977 4.28e-07 ***
## abs(phi)     0.02959    0.04499   0.658    0.514
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.06759 on 42 degrees of freedom
## Multiple R-squared:  0.01019,   Adjusted R-squared:  -0.01337
## F-statistic: 0.4325 on 1 and 42 DF,  p-value: 0.5144
```

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

```
##
## Call:
## lm(formula = Delta_under ~ abs(corrected_phi), data = gram)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.08416 -0.04707 -0.02479  0.04093  0.15041
##
## Coefficients:
##                  Estimate Std. Error t value Pr(>|t|)
## (Intercept)       0.06863    0.01360   5.048  9.1e-06 ***
## abs(corrected_phi) 0.11874    0.06488   1.830   0.0743 .
```

```
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.06538 on 42 degrees of freedom
## Multiple R-squared:  0.07385,    Adjusted R-squared:  0.0518
## F-statistic: 3.349 on 1 and 42 DF,  p-value: 0.07435
```

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

```
## [1] -108.2768
```

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

```
## [1] -111.2017
```

```r
#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

```r
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
## -0.090293 -0.008951  0.011975  0.018441  0.019683
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) -0.017842   0.007019  -2.542   0.0148 *
## abs(phi)    -0.002075   0.017905  -0.116   0.9083
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.02639 on 42 degrees of freedom
## Multiple R-squared:  0.0003195, Adjusted R-squared:  -0.02348
## F-statistic: 0.01342 on 1 and 42 DF,  p-value: 0.9083
```

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

```
##
## Call:
## lm(formula = Delta_over ~ abs(corrected_phi), data = wals)
##
## Residuals:
##       Min        1Q    Median        3Q       Max
## -0.090947 -0.009303  0.012863  0.017843  0.020939
##
```

6

```
## Coefficients:
##                   Estimate Std. Error t value Pr(>|t|)
## (Intercept)       -0.017463   0.006103  -2.861  0.00655 **
## abs(corrected_phi) -0.005431   0.023966  -0.227  0.82183
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.02638 on 42 degrees of freedom
## Multiple R-squared:  0.001221,   Adjusted R-squared:  -0.02256
## F-statistic: 0.05135 on 1 and 42 DF,  p-value: 0.8218
```

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

```
## [1] -191.0436
```

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

```
## [1] -191.0833
```

```
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
## -0.049972 -0.010194  0.006076  0.013586  0.022462
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -0.011715   0.003751  -3.123  0.00323 **
## abs(phi)    -0.020381   0.012599  -1.618  0.11323
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.01893 on 42 degrees of freedom
## Multiple R-squared:  0.05865,    Adjusted R-squared:  0.03624
## F-statistic: 2.617 on 1 and 42 DF,  p-value: 0.1132
```

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

```
##
## Call:
## lm(formula = Delta_over ~ abs(corrected_phi), data = gram)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.055741 -0.006846  0.006055  0.015018  0.016961
##
## Coefficients:
##                   Estimate Std. Error t value Pr(>|t|)
## (Intercept)       -0.014741   0.004052  -3.638 0.000745 ***
## abs(corrected_phi) -0.006312   0.019337  -0.326 0.745705
```

```
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.01949 on 42 degrees of freedom
## Multiple R-squared:  0.002531,   Adjusted R-squared:  -0.02122
## F-statistic: 0.1066 on 1 and 42 DF,  p-value: 0.7457
```

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

```
## [1] -220.2782
```

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

```
## [1] -217.7303
```