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

Statistical analyses

Preliminaries

```r
# required packages
require(tidyverse)
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$dataset == "WALS" & infl$inflpoint > 1, ]$inflpoint))
ip_grambank <- round(mean(infl[infl$dataset == "Grambank" & infl$inflpoint > 1, ]$inflpoint))

# 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
## -0.07397 -0.03576 -0.01483  0.03319  0.11248
##
## Coefficients:
##                   Estimate Std. Error t value Pr(>|t|)
## (Intercept)        0.04331    0.01565   2.768  0.00851 **
## statusinteracting  0.05335    0.02555   2.088  0.04322 *
## statusunknown      0.02231    0.01711   1.304  0.19970
## abs(phi)          -0.05044    0.04563  -1.105  0.27558
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.04884 on 40 degrees of freedom
## Multiple R-squared:  0.1044, Adjusted R-squared:  0.03726
## F-statistic: 1.555 on 3 and 40 DF,  p-value: 0.2154
```

```
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
## -0.075847 -0.000379  0.007408  0.011577  0.023463
##
## Coefficients:
##                    Estimate Std. Error t value Pr(>|t|)
## (Intercept)       -0.006151   0.007239  -0.850   0.4005
## statusinteracting -0.002596   0.011820  -0.220   0.8273
## statusunknown     -0.013785   0.007915  -1.742   0.0893 .
## abs(phi)          -0.008039   0.021106  -0.381   0.7053
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.02259 on 40 degrees of freedom
## Multiple R-squared:  0.07848,    Adjusted R-squared:  0.009369
## F-statistic: 1.136 on 3 and 40 DF,  p-value: 0.3463
```

```
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
## -0.05851 -0.03812 -0.01825  0.02337  0.11469
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.046172   0.013387   3.449  0.00129 **
## abs(phi)    0.008696   0.034147   0.255  0.80022
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.05033 on 42 degrees of freedom
## Multiple R-squared:  0.001542,   Adjusted R-squared:  -0.02223
## F-statistic: 0.06486 on 1 and 42 DF,  p-value: 0.8002
```

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

```
##
## Call:
## lm(formula = Delta_under ~ abs(corrected_phi), data = wals)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.06297 -0.03466 -0.01258  0.02747  0.11582
##
## Coefficients:
##                    Estimate Std. Error t value Pr(>|t|)
## (Intercept)         0.02852    0.01089   2.620   0.0122 *
## abs(corrected_phi)  0.10592    0.04274   2.478   0.0173 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.04704 on 42 degrees of freedom
## Multiple R-squared:  0.1276, Adjusted R-squared:  0.1068
## F-statistic: 6.141 on 1 and 42 DF,  p-value: 0.01731
```

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

```
## [1] -134.2299
```

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

```
## [1] -140.1662
```

```
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
```

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

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

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

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

```
#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
## -0.084237 -0.004278  0.010180  0.014321  0.017072
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -0.013378   0.006104  -2.192    0.034 *
## abs(phi)    -0.004163   0.015570  -0.267    0.790
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.02295 on 42 degrees of freedom
## Multiple R-squared:  0.001699,   Adjusted R-squared:  -0.02207
## F-statistic: 0.07148 on 1 and 42 DF,  p-value: 0.7905
```

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

```
##
## Call:
## lm(formula = Delta_over ~ abs(corrected_phi), data = wals)
##
## Residuals:
##       Min        1Q    Median        3Q       Max
## -0.085176 -0.003947  0.010331  0.014126  0.016952
##
## Coefficients:
##                    Estimate Std. Error t value Pr(>|t|)
## (Intercept)        -0.01376    0.00531  -2.591   0.0131 *
## abs(corrected_phi) -0.00499    0.02085  -0.239   0.8120
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.02295 on 42 degrees of freedom
## Multiple R-squared:  0.001362,   Adjusted R-squared:  -0.02242
## F-statistic: 0.05727 on 1 and 42 DF,  p-value: 0.812
```

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

```
## [1] -203.3416
```

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

```
## [1] -203.3267
```

```
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
```