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

Statistical analyses

Preliminaries

```
# 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
```

```
require(broom)
```

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

```
require(pixiedust)
```

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

```
require(emmeans)
```

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

```
# 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
```

```
# 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
```

```r
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^-)

```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
## -0.07680 -0.03759 -0.01338  0.03499  0.12071
##
## Coefficients:
##                   Estimate Std. Error t value Pr(>|t|)
## (Intercept)        0.04764    0.01507   3.161  0.00300 **
## statusinteracting  0.06369    0.02308   2.760  0.00868 **
## statusunknown      0.01564    0.01701   0.919  0.36347
## abs(phi)          -0.07062    0.04365  -1.618  0.11350
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.04725 on 40 degrees of freedom
## Multiple R-squared:  0.1618, Adjusted R-squared:  0.09897
## F-statistic: 2.574 on 3 and 40 DF,  p-value: 0.06734
```

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

```
##
## Call:
## lm(formula = Delta_under ~ status + abs(phi), data = gram)
##
## Residuals:
##       Min        1Q    Median        3Q       Max
```

```
## -0.104664 -0.049021  0.000074  0.042763  0.125229
##
## Coefficients:
##                   Estimate Std. Error t value Pr(>|t|)
## (Intercept)        0.05278    0.01654   3.192  0.00275 **
## statusinteracting  0.08557    0.03485   2.455  0.01852 *
## statusunknown      0.05536    0.02300   2.407  0.02081 *
## abs(phi)          -0.03944    0.04906  -0.804  0.42611
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.0634 on 40 degrees of freedom
## Multiple R-squared:  0.1707, Adjusted R-squared:  0.1085
## F-statistic: 2.744 on 3 and 40 DF,  p-value: 0.0556
```

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
## -0.077798 -0.002361  0.007683  0.013729  0.021706
##
## Coefficients:
##                    Estimate Std. Error t value Pr(>|t|)
## (Intercept)       -0.007371   0.007266  -1.014    0.316
## statusinteracting -0.006961   0.011124  -0.626    0.535
## statusunknown     -0.013303   0.008200  -1.622    0.113
## abs(phi)          -0.002352   0.021039  -0.112    0.912
##
## Residual standard error: 0.02278 on 40 degrees of freedom
## Multiple R-squared:  0.06333,    Adjusted R-squared:  -0.006916
## F-statistic: 0.9016 on 3 and 40 DF,  p-value: 0.4489
```

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

```
##
## Call:
## lm(formula = Delta_over ~ status + abs(phi), data = gram)
##
## Residuals:
##       Min        1Q    Median        3Q       Max
## -0.048495 -0.012389  0.006823  0.010652  0.023535
##
## Coefficients:
```

```
##                    Estimate Std. Error t value Pr(>|t|)
## (Intercept)       -0.008684   0.005001  -1.736   0.0902 .
## statusinteracting -0.003723   0.010540  -0.353   0.7257
## statusunknown     -0.006643   0.006957  -0.955   0.3454
## abs(phi)          -0.015469   0.014836  -1.043   0.3034
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.01917 on 40 degrees of freedom
## Multiple R-squared:  0.08033,    Adjusted R-squared:  0.01136
## F-statistic: 1.165 on 3 and 40 DF,  p-value: 0.3352
```

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
```