Appendix 1

1. Predicting aspectual choice in authentic samples with:

1.1. Unigram general aspectual preferences:

```
> GLM_auth_ug<-glm(formula = aspect ~ ug_pref, family = "binomial", data = a_all)
> summary(GLM_auth_ug)

Call:
glm(formula = aspect ~ ug_pref, family = "binomial", data = a_all)

Deviance Residuals:
   Min      1Q  Median      3Q     Max
-1.713  -1.319   1.042   1.042   1.448

Coefficients:
                  Estimate Std. Error z value Pr(>|z|)
(Intercept)        -0.6174     0.1791  -3.448 0.000565 ***
ug_prefimpf/pf      1.8214     0.4987   3.652 0.000260 ***
ug_prefpf           0.9446     0.1966   4.804 1.56e-06 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 1082.7  on 785  degrees of freedom
Residual deviance: 1052.8  on 783  degrees of freedom
AIC: 1058.8

Number of Fisher Scoring iterations: 4

Analysis of Deviance Table

Model: binomial, link: logit

Response: aspect
Terms added sequentially (first to last)


           Df Deviance Resid. Df Resid. Dev  Pr(>Chi)
NULL                         785     1082.7
ug_pref     2   29.882       783     1052.8 3.244e-07 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

> exp(cbind(OR = coef(GLM_auth_ug), confint(GLM_auth_ug))) #checking odds ratios
Waiting for profiling to be done...
                       OR     2.5 %     97.5 %
(Intercept)     0.5393258 0.3769933  0.7620924
ug_prefimpf/pf 6.1805556 2.4501467 17.8512002
ug_prefpf       2.5716794 1.7571324  3.8035037

> LRM_auth_ug = lrm(aspect ~ ug_pref, data = a_all, x=T, y=T)
> LRM_auth_ug

Logistic Regression Model

lrm(formula = aspect ~ ug_pref, data = a_all, x = T, y = T)
                       Model Likelihood     Discrimination    Rank Discrim.
```

doi: https://doi.org/10.1075/target.15155.szy.additional

```
                         Ratio Test            Indexes          Indexes
Obs           786    LR chi2      29.88    R2        0.050    C         0.579
 impf         356    d.f.             2    g         0.328    Dxy       0.158
 pf           430    Pr(> chi2) <0.0001    gr        1.389    gamma     0.447
max |deriv| 1e-11                          gp        0.079    tau-a     0.079
                                           Brier     0.238

                Coef    S.E.   Wald Z Pr(>|Z|)
Intercept       -0.6174 0.1791 -3.45  0.0006
ug_pref=impf/pf  1.8214 0.4987  3.65  0.0003
ug_pref=pf       0.9446 0.1966  4.80  <0.0001
```

1.2. Unigram legal aspectual preferences:

```
> GLM_auth_ul<-glm(formula = aspect ~ ul_pref, family = "binomial", data = a_all)

Call:
glm(formula = aspect ~ ul_pref, family = "binomial", data = a_all)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-1.4894  -1.0033   0.8947   0.8947   1.3621

Coefficients:
                  Estimate Std. Error z value Pr(>|z|)
(Intercept)        -0.4243     0.1160  -3.658 0.000254 ***
ul_prefimpf/pf   0.6184     0.2385   2.593 0.009506 **
ul_prefpf        1.1332     0.1590   7.125 1.04e-12 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 1082.7  on 785  degrees of freedom
Residual deviance: 1029.9  on 783  degrees of freedom
AIC: 1035.9
Number of Fisher Scoring iterations: 4

Analysis of Deviance Table

Model: binomial, link: logit

Response: aspect

Terms added sequentially (first to last)

           Df Deviance Resid. Df Resid. Dev  Pr(>Chi)
NULL                         785     1082.7
ul_pref     2   52.718       783     1029.9 3.568e-12 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

> exp(cbind(OR = coef(GLM_auth_ul), confint(GLM_auth_ul)))
Waiting for profiling to be done...
                       OR     2.5 %    97.5 %
(Intercept)    0.6542553 0.5201643 0.8199479
ul_prefimpf/pf 1.8559814 1.1647804 2.9721320
ul_prefpf      3.1054330 2.2783597 4.2511527
```

```
> LRM_auth_ul = lrm(aspect ~ ul_pref, data = a_all, x=T, y=T)
> LRM_auth_ul

Logistic Regression Model

lrm(formula = aspect ~ ul_pref, data = a_all, x = T, y = T)
                       Model Likelihood     Discrimination    Rank Discrim.
                          Ratio Test            Indexes          Indexes
Obs           786    LR chi2      52.72    R2        0.087    C       0.635
 impf         356    d.f.             2    g         0.554    Dxy     0.270
 pf           430    Pr(> chi2) <0.0001    gr        1.740    gamma   0.433
max |deriv| 5e-13                          gp        0.134    tau-a   0.134
                                           Brier     0.231

                Coef    S.E.   Wald Z Pr(>|Z|)
Intercept       -0.4243 0.1160 -3.66  0.0003
ul_pref=impf/pf  0.6184 0.2385  2.59  0.0095
ul_pref=pf       1.1332 0.1590  7.13  <0.0001
```

1.3. Chunk general apsectual preferences:

```
> GLM_auth_cg<-glm(formula = aspect ~ cg_pref, family = "binomial", data = a_all)

Call:
glm(formula = aspect ~ cg_pref, family = "binomial", data = a_all)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-1.7379  -0.9125   0.7065   0.7065   1.7663

Coefficients:
                Estimate Std. Error z value Pr(>|z|)
(Intercept)      -1.3241     0.2089  -6.337 2.34e-10 ***
cg_prefimpf/pf   0.6631     0.2507   2.645  0.00817 **
cg_prefna        0.2254     1.1735   0.192  0.84765
cg_prefpf        2.5846     0.2403  10.754  < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 1082.65  on 785  degrees of freedom
Residual deviance:  879.08  on 782  degrees of freedom
AIC: 887.08

Number of Fisher Scoring iterations: 4

Analysis of Deviance Table

Model: binomial, link: logit

Response: aspect

Terms added sequentially (first to last)

         Df Deviance Resid. Df Resid. Dev  Pr(>Chi)
NULL                       785    1082.65
```

```
cg_pref     3   203.56       782     879.08 < 2.2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

```
> exp(cbind(OR = coef(GLM_auth_cg), confint(GLM_auth_cg)))
Waiting for profiling to be done...
                       OR      2.5 %     97.5 %
(Intercept)     0.266055 0.17352375  0.3948626
cg_prefimpf/pf  1.940726 1.19780724  3.2089022
cg_prefna       1.252874 0.06075197 10.2004476
cg_prefpf      13.258431 8.38094763 21.5511366
```

```
> LRM_auth_cg = lrm(aspect ~ cg_pref, data = a_all, x=T, y=T)
> LRM_auth_cg

Logistic Regression Model

lrm(formula = aspect ~ cg_pref, data = a_all, x = T, y = T)
                       Model Likelihood     Discrimination    Rank Discrim.
                          Ratio Test            Indexes          Indexes
Obs           786    LR chi2     203.57    R2       0.305    C       0.759
 impf         356    d.f.             3    g        1.155    Dxy     0.519
 pf           430    Pr(> chi2) <0.0001    gr       3.173    gamma   0.731
max |deriv| 7e-10                          gp       0.257    tau-a   0.257
                                           Brier    0.187

                Coef    S.E.   Wald Z Pr(>|Z|)
Intercept       -1.3241 0.2089 -6.34  <0.0001
cg_pref=impf/pf  0.6631 0.2507  2.64  0.0082
cg_pref=na       0.2254 1.1735  0.19  0.8477
cg_pref=pf       2.5846 0.2403 10.75  <0.0001
```

1.4. Chunk legal aspectual preferences:

```
> GLM_auth_cl<-glm(formula = aspect ~ cl_pref, family = “binomial”, data = a_all)

Call:
glm(formula = aspect ~ cl_pref, family = “binomial”, data = a_all)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-2.0237  -1.1433   0.5256   1.2120   1.8676

Coefficients:
                Estimate Std. Error z value Pr(>|z|)
(Intercept)      -1.5518     0.2202  -7.048 1.81e-12 ***
cl_prefimpf/pf   1.4709     0.2435   6.042 1.53e-09 ***
cl_prefna        2.3562     0.3655   6.446 1.15e-10 ***
cl_prefpf        3.4614     0.2991  11.571  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 1082.65  on 785  degrees of freedom
Residual deviance:  881.17  on 782  degrees of freedom
AIC: 889.17

Number of Fisher Scoring iterations: 4
```

```
Analysis of Deviance Table

Model: binomial, link: logit

Response: aspect

Terms added sequentially (first to last)

          Df Deviance Resid. Df Resid. Dev  Pr(>Chi)
NULL                        785    1082.65
cl_pref  3   201.48       782     881.17 < 2.2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

```
> exp(cbind(OR = coef(GLM_auth_cl), confint(GLM_auth_cl)))
Waiting for profiling to be done...
                        OR      2.5 %     97.5 %
(Intercept)      0.2118644  0.1345898  0.320251
cl_prefimpf/pf   4.3531606  2.7417862  7.144554
cl_prefna       10.5505882  5.2484430 22.105160
cl_prefpf       31.8600000 18.0607735 58.486759
```

```
> LRM_auth_cl = lrm(aspect ~ cl_pref, data = a_all, x=T, y=T)
> LRM_auth_cl

Logistic Regression Model

lrm(formula = aspect ~ cl_pref, data = a_all, x = T, y = T)
                       Model Likelihood     Discrimination    Rank Discrim.
                          Ratio Test            Indexes          Indexes
Obs           786    LR chi2     201.48    R2       0.302    C       0.761
 impf         356    d.f.             3    g        1.282    Dxy     0.521
 pf           430    Pr(> chi2) <0.0001    gr       3.604    gamma   0.726
max |deriv| 1e-12                          gp       0.259    tau-a   0.259
                                           Brier    0.190

                 Coef    S.E.   Wald Z Pr(>|Z|)
Intercept       -1.5518 0.2202 -7.05  <0.0001
cl_pref=impf/pf  1.4709 0.2435  6.04  <0.0001
cl_pref=na       2.3562 0.3655  6.45  <0.0001
cl_pref=pf       3.4614 0.2991 11.57  <0.0001
```

2. Predicting aspectual choice in translated samples with:

2.1. Unigram general aspectual preferences:

```
> GLM_trans_ug<-glm(formula = aspect ~ ug_pref, family = “binomial”, data = t_all)

Call:
glm(formula = aspect ~ ug_pref, family = “binomial”, data = all_trans)

Deviance Residuals:
   Min      1Q  Median      3Q     Max
-1.266  -1.266   1.091   1.091   1.356

Coefficients:
                  Estimate Std. Error z value Pr(>|z|)
(Intercept)        -0.4106     0.1600  -2.567 0.010267 *
ug_prefimpf/pf   0.6047     0.3948   1.532 0.125551
ug_prefpf        0.6168     0.1846   3.342 0.000833 ***
```

```
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 929.67  on 670  degrees of freedom
Residual deviance: 918.14  on 668  degrees of freedom
AIC: 924.14

Number of Fisher Scoring iterations: 4

Analysis of Deviance Table

Model: binomial, link: logit

Response: aspect

Terms added sequentially (first to last)

           Df Deviance Resid. Df Resid. Dev Pr(>Chi)
NULL                         670     929.67
ug_pref     2   11.523       668     918.14 0.003147 **
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

> exp(cbind(OR = coef(GLM_trans_ug), confint(GLM_trans_ug)))
Waiting for profiling to be done...
                       OR     2.5 %    97.5 %
(Intercept)     0.6632653 0.4828306 0.9050807
ug_prefimpf/pf 1.8307692 0.8458902 4.0214630
ug_prefpf       1.8529116 1.2933206 2.6686963

> LRM_trans_ug = lrm(aspect ~ ug_pref, data = t_all, x=T, y=T)
> LRM_trans_ug

Logistic Regression Model

lrm(formula = aspect ~ unigram_gen, data = t_all, x = T, y = T)
                      Model Likelihood     Discrimination    Rank Discrim.
                         Ratio Test            Indexes          Indexes
Obs            671    LR chi2     11.52    R2       0.023    C       0.556
 impf          326    d.f.            2    g        0.228    Dxy     0.113
 pf            345    Pr(> chi2) 0.0031    gr       1.256    gamma   0.255
max |deriv| 3e-11                         gp       0.056    tau-a   0.056
                                          Brier    0.246

                     Coef    S.E.   Wald Z Pr(>|Z|)
Intercept           -0.4106 0.1600 -2.57  0.0103
ug_pref=impf/pf  0.6047 0.3948  1.53  0.1256
ug_pref=pf       0.6168 0.1846  3.34  0.0008
```

2.2. Unigram legal aspectual preferences:

```
> GLM_trans_ul<-glm(formula = aspect ~ ul_pref, family = “binomial”, data = t_all)

Call:
glm(formula = aspect ~ unigram_leg, family = “binomial”, data = all_trans)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
```

```
-1.3815  -1.2018   0.9862   0.9862   1.3830

Coefficients:
                   Estimate Std. Error z value Pr(>|z|)
(Intercept)         -0.4713     0.1322  -3.567 0.000361 ***
unigram_legimpf/pf   0.5285     0.2358   2.242 0.024990 *
unigram_legna      -13.0947   378.5929  -0.035 0.972408
unigram_legpf        0.9393     0.1749   5.372  7.8e-08 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 929.67  on 670  degrees of freedom
Residual deviance: 897.11  on 667  degrees of freedom
AIC: 905.11

Number of Fisher Scoring iterations: 12

Analysis of Deviance Table

Model: binomial, link: logit

Response: aspect

Terms added sequentially (first to last)

            Df Deviance Resid. Df Resid. Dev  Pr(>Chi)
NULL                          670     929.67
unigram_leg  3   32.556       667     897.11 3.995e-07 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
> exp(cbind(OR = coef(GLM_trans_ul), confint(GLM_trans_ul)))
Waiting for profiling to be done...
                          OR     2.5 %       97.5 %
(Intercept)     6.241611e-01 0.4803089 8.068968e-01
ul_prefimpf/pf 1.696395e+00 1.0692377 2.698365e+00
ul_prefna       2.056059e-06        NA 1.329395e+20
ul_prefpf       2.558273e+00 1.8197124 3.613253e+00
Warning messages:
1: glm.fit: fitted probabilities numerically 0 or 1 occurred

> LRM_trans_ul = lrm(aspect ~ ul_pref, data = t_all, x=T, y=T)
> LRM_trans_ul

Logistic Regression Model

lrm(formula = aspect ~ unigram_leg, data = t_all, x = T, y = T)
                       Model Likelihood     Discrimination    Rank Discrim.
                          Ratio Test            Indexes          Indexes
Obs           671    LR chi2      32.55    R2        0.063    C       0.613
 impf         326    d.f.             3    g         0.490    Dxy     0.226
 pf           345    Pr(> chi2) <0.0001    gr        1.633    gamma   0.356
max |deriv| 0.001                          gp        0.113    tau-a   0.113
                                           Brier     0.238

                    Coef    S.E.    Wald Z Pr(>|Z|)
Intercept           -0.4713  0.1322 -3.57  0.0004
```

```
ul_pref=impf/pf  0.5285  0.2358  2.24  0.0250
ul_pref=na      -6.7317 25.9378 -0.26  0.7952
ul_pref=pf       0.9393  0.1749  5.37  <0.0001
```

2.3. Chunk general apsectual preferences:

```
> GLM_trans_cg<-glm(formula = aspect ~ cg_pref, family = "binomial", data = t_all)

Call:
glm(formula = aspect ~ chunk_gen, family = "binomial", data = all_trans)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-1.6263  -0.8954   0.7873   0.7873   1.7181

Coefficients:
                 Estimate Std. Error z value Pr(>|z|)
(Intercept)       -1.2164     0.2324  -5.234 1.66e-07 ***
chunk_genimpf/pf   0.5095     0.2734   1.863   0.0624 .
chunk_genna        0.3001     0.6356   0.472   0.6368
chunk_genpf        2.2290     0.2633   8.466  < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 929.67  on 670  degrees of freedom
Residual deviance: 793.47  on 667  degrees of freedom
AIC: 801.47

Number of Fisher Scoring iterations: 4

Analysis of Deviance Table

Model: binomial, link: logit

Response: aspect

Terms added sequentially (first to last)


          Df Deviance Resid. Df Resid. Dev  Pr(>Chi)
NULL                        670     929.67
chunk_gen  3   136.19       667     793.47 < 2.2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

> exp(cbind(OR = coef(GLM_trans_cg), confint(GLM_trans_cg)))
Waiting for profiling to be done...
                       OR     2.5 %     97.5 %
(Intercept)     0.2962963 0.1839782  0.4595076
cg_prefimpf/pf 1.6643836 0.9843430  2.8851652
cg_prefna       1.3500000 0.3459708  4.4462225
cg_prefpf       9.2907303 5.6239853 15.8395027

> LRM_trans_cg = lrm(aspect ~ cg_pref, data = t_all, x=T, y=T)
> LRM_trans_cg

Logistic Regression Model
```

```
lrm(formula = aspect ~ chunk_gen, data = t_all, x = T, y = T)
                      Model Likelihood     Discrimination    Rank Discrim.
                         Ratio Test           Indexes          Indexes
Obs           671    LR chi2     136.19    R2        0.245    C        0.730
 impf         326    d.f.             3    g         1.002    Dxy      0.460
 pf           345    Pr(> chi2) <0.0001    gr        2.723    gamma    0.662
max |deriv| 1e-10                          gp        0.230    tau-a    0.230
                                           Brier     0.201

                  Coef    S.E.   Wald Z Pr(>|Z|)
Intercept         -1.2164 0.2324 -5.23  <0.0001
cg_pref=impf/pf   0.5095 0.2734  1.86   0.0624
cg_pref=na        0.3001 0.6356  0.47   0.6368
cg_pref=pf        2.2290 0.2633  8.47   <0.0001
```

2.4. Chunk legal aspectual preferences:

```
> GLM_trans_cl<-glm(formula = aspect ~ cl_pref, family = "binomial", data = t_all)

Call:
glm(formula = aspect ~ chunk_leg, family = "binomial", data = all_trans)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-1.6861  -1.2031   0.7433   1.1520   1.3824

Coefficients:
                Estimate Std. Error z value Pr(>|z|)
(Intercept)      -0.47000    0.25495  -1.844   0.0653 .
chunk_legimpf/pf  0.53022    0.27998   1.894   0.0583 .
chunk_legna       0.09932    0.29440   0.337   0.7358
chunk_legpf       1.61514    0.33478   4.824  1.4e-06 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 929.67  on 670  degrees of freedom
Residual deviance: 887.40  on 667  degrees of freedom
AIC: 895.4

Number of Fisher Scoring iterations: 4
Analysis of Deviance Table

Model: binomial, link: logit

Response: aspect

Terms added sequentially (first to last)

          Df Deviance Resid. Df Resid. Dev Pr(>Chi)
NULL                        670     929.67
chunk_leg  3   42.268       667     887.40 3.52e-09 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

> GLM_trans_cl<-glm(formula = aspect ~ cl_pref, family = "binomial", data = t_all)
> exp(cbind(OR = coef(GLM_trans_cl), confint(GLM_trans_cl)))
Waiting for profiling to be done...
                    OR     2.5 %   97.5 %
```

```
(Intercept)     0.625000 0.3745490 1.022934
cl_prefimpf/pf 1.699310 0.9875611 2.971739
cl_prefna       1.104425 0.6231789 1.983543
cl_prefpf       5.028571 2.6365091 9.826942
```

```
> LRM_trans_cl = lrm(aspect ~ cl_pref, data = t_all, x=T, y=T)
> LRM_trans_cl
```

```
Logistic Regression Model

lrm(formula = aspect ~ chunk_leg, data = t_all, x = T, y = T)
                      Model Likelihood     Discrimination    Rank Discrim.
                         Ratio Test            Indexes          Indexes
Obs           671    LR chi2      42.27    R2        0.081    C         0.624
 impf         326    d.f.             3    g         0.532    Dxy       0.249
 pf           345    Pr(> chi2) <0.0001    gr        1.702    gamma     0.359
max |deriv| 3e-11                          gp        0.124    tau-a     0.124
                                           Brier     0.235

                  Coef    S.E.   Wald Z Pr(>|Z|)
Intercept         -0.4700 0.2550 -1.84  0.0653
cl_pref=impf/pf  0.5302 0.2800  1.89  0.0583
cl_pref=na       0.0993 0.2944  0.34  0.7358
cl_pref=pf       1.6151 0.3348  4.82  <0.0001
```