Appendix E Statistical modelling in Chapter 3

E.1. Introduction

This appendix provides the details of our multifactorial (logistic) regression modelling, the results of which were presented in Section 3.4.2 to 3.4.4. We start with the dataset, variable definition and coding, and then provide full summaries of the models along with details on model selection and model quality.

As mentioned in the text, all the analyses were done in R. The code and data can be found at the following link:

https://osf.io/pvmbw/?view_only=43f05df2f2954951a5945ea4ac8fad5f

E.2. Dataset and variable coding

The goal of our modelling was to investigate the effect of a set of functional factors (animacy, argument role, argument realisation and topicality) in differential indexing. More precisely, we wanted to estimate the contribution of each factor in determining the choice of indexing strategy in the case of alternating verbs. Hence, the data is limited to the alternating subset of the corpus, including only the following four major marking strategies: /e/-series, /o/-series, /a/-series and zero.

E.2.1. Predicting variables

Choosing between Topic persistence (TP) and Referential distance (RD) as measures of topicality:

These measures are highly correlated:

```
lm(formula = ArgTP1 ~ ArgRD1, data = AltDS)

Residuals:
    Min      1Q  Median      3Q     Max 
-4.0809 -1.9809 -0.7943  1.9191  6.0921 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept)  4.25391    0.19617  21.685   <2e-16 ***
ArgRD1      -0.17298    0.01817  -9.521   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 2.614 on 316 degrees of freedom
Multiple R-squared:  0.2229,     Adjusted R-squared:  0.2204 
F-statistic: 90.64 on 1 and 316 DF,  p-value: < 2.2e-16
```

We compared goodness-of-fit to select the best predictor:

```
multinom(formula = choice ~
ArgTP1, data = AltDS)

Coefficients:
  (Intercept)      ArgTP1
1  -1.5369436  0.064471241
2  -0.5920302 -0.489911647
3  -2.1737311  0.009409013

Std. Errors:
  (Intercept)     ArgTP1
1   0.2455880 0.05117558
2   0.2045975 0.10819499
3   0.3317169 0.07424120

Residual Deviance: 637.1292
AIC: 649.1292
-> R2 = 0.12
```

```
multinom(formula = choice ~
ArgRD1, data = AltDS)

Coefficients:
  (Intercept)      ArgRD1
1   -1.150103 -0.02595372
2   -2.146889  0.07385478
3   -1.851602 -0.05405053

Std. Errors:
  (Intercept)     ArgRD1
1   0.1969863 0.02150812
2   0.2722856 0.01933959
3   0.2707848 0.03579558

Residual Deviance: 655.1943
AIC: 667.19
-> R2 = 0.80
```

The predicting factors investigated in our analysis were defined as follows:
- Argument role as a binary variable (S vs. P) → ArgType
- Animacy as a binary variable (animate vs. inanimate) → ArgAnim
- Argument realisation as a nominal variable with three levels (NP, Pro and Null) → ArgRlz
- Topic persistence as a discrete variable with values spanning from zero to 10 → ArgTP1

Our nominal predicting variables were sum-coded (or centred), as follows:

```
ArgType:          ArgAnime:          ArgRlz:
       [,1]              [,1]               [,1]   [,2]
S        1        Anim     1         Pro      1      0
P       -1        Inan    -1         NP       0      1
                                     Null    -1     -1
```

Given these definitions, the final dataset included 318 data points. Note that the ratios reported in the main text are calculated for the whole dataset (i.e., the total of 797 tokens, of which 365 are alternating verbs), whereas the coefficient estimates are drawn from the multifactorial modelling results presented in this appendix.

E.2.2. Response variable

In the multinomial model, the response variable (named "choice") codes the indexing strategy on 4 levels, with "zero" as the baseline:

```
0  →    zero
1  →    /e/-series (named "GEN" in the data frame)
2  →    /o/-series (named "LOC" in the data frame)
3  →    /a/-series (named "PAT" in the data frame)
```

In the binomial models, fitted for each strategy separately, the response variable is always a binary variable coding presence (=1/success) vs. absence (=0/failure) of the given strategy. Note that the prefixes were named as 'GEN', 'LOC' and 'PAT' (instead of /e/-series, /o/-series, and /a/-series, respectively) for coding convenience.

E.3. Models

To study the effect of our four factors, we fitted a multinomial model (Section E.3.1) and a series of four binary fixed-effects models (Section E.3.2) to the whole dataset (named 'AltDS'), without taking verb lemmas into account. These are the models from which we reported the coefficient estimates for different effects in Section 3.4.2 and Section 3.4.3 of the main text.

Next, aiming to estimate the role of lexical stipulation (Section 4.4 in the main text), we first fitted a second series of fixed-effect binary models (Section E.3.3) to subsets of the data limited to the verbs whose alternation profile includes the prefix under investigation; that is, is coded as the response variable. Finally, we fitted a series of mixed-effects models with the verbal lemma as random intercept (Section E.3.4). We compare the quality of the fit of our three series of binary models in Section E.3..5.

E.3.1. Multinomial model – fitted to the whole dataset (Sections 3.4.2 & 3.4.3)

```
318 tokens
Distribution:
       /e/ /o/ /a/   Z
     0   0   0   0 196
     1  53   0   0   0
     2   0  46   0   0
     3   0   0  23   0
```

Model selection:

We report below only the results of the most relevant model comparisons using Likelihood ratio tests of Multinomial Models.

```
                                Model Resid. df Resid. Dev   Test         Df LR stat.    Pr(Chi)
1          ArgAnim + ArgTP1 + ArgType        942   516.0940
2 ArgType + ArgAnim + ArgRlz + ArgTP1        936   497.5383 1 vs 2        6  18.5557     0.004983645

1          ArgAnim + ArgRlz + ArgType        939   501.1369
2 ArgType + ArgAnim + ArgRlz + ArgTP1        936   497.5383 1 vs 2        3 3.598616     0.3081953

1          ArgAnim + ArgRlz                  942   567.9497
2 ArgAnim + ArgRlz + ArgTP1                  939   557.4552 1 vs 2        3 10.49449     0.01479835

1          ArgType + ArgRlz                  942   563.8842
2 ArgType + ArgRlz + ArgTP1                  939   553.2466 1 vs 2        3 10.63759     0.013856

1 ArgType + ArgRlz + ArgTP1                  939   553.2466
2 ArgRlz + ArgType * ArgTP1                  936   546.5991 1 vs 2        3 6.647497     0.08402358

1           ArgRlz + ArgType * ArgTP1        936   546.5991
2 ArgAnim + ArgRlz + ArgType * ArgTP1        933   495.5851 1 vs 2        3 51.01402     4.858558e-11

1 ArgType + ArgAnim + ArgRlz + ArgTP1        936   497.5383
2 ArgType * ArgAnim + ArgRlz + ArgTP1        933   497.4341 1 vs 2        3 0.1041331    0.9913369
```

Summary of results

```
Call:
multinom(formula = choice ~ ArgAnim + ArgRlz + ArgType * ArgTP1, data = AltDS)
```

```
Coefficients:
  (Intercept)   ArgAnim1    ArgRlz1    ArgRlz2   ArgType1      ArgTP1 ArgType1:ArgTP1
1   -1.654055  0.7767583 -0.7503987  0.3079579 -0.1007462  0.01662436     -0.06865582
2   -8.995567 -0.3790691 -0.3005077 -0.1344887 -8.5379892 -0.30970969     -0.11383577
3  -10.221565  8.9382960 -1.1934081 -0.1738070 -1.1483932 -0.05561162     -0.12619068

Std. Errors:
  (Intercept)  ArgAnim1   ArgRlz1   ArgRlz2  ArgType1     ArgTP1   ArgType1:ArgTP1
1   0.2799792 0.2656414 0.3037171 0.2482772 0.3073998 0.07494490      0.07225559
2   0.2038896 0.3638613 0.4987661 0.3179002 0.2038896 0.06783363      0.06783366
3   0.2477353 0.2477355 0.5614181 0.4629413 0.4508107 0.09978019      0.10202008

Residual Deviance: 495.5851
AIC: 537.
```

Goodness-of-fit

Null Model:

```
MN_M0 <-multinom(choice~1, data=AltDS)
```

R2

```
r.squaredLR(MN_M1, MN_M0)
[1] 0.4371003
attr(,"adj.r.squared")
[1] 0.4958427
```

Observed/predicted contingency tables:

	0	1	2	3
0	185	0	10	1
1	47	0	1	5
2	31	0	14	1
3	12	0	0	11

-> Accuracy = 66.04

	0	1	2	3
0	196	0	0	0
1	53	0	0	0
2	46	0	0	0
3	23	0	0	0

-> Accuracy = 61.64

P-values (using z-test):

	Estimate	Std. Error	z value	Pr(>|z|)	
1:(Intercept)	-1.654055	0.279979	-5.9078	3.468e-09	***
1:ArgAnim1	0.776758	0.265641	2.9241	0.003455	**
1:ArgRlz1	-0.750399	0.303717	-2.4707	0.013484	*
1:ArgRlz2	0.307958	0.248277	1.2404	0.214835	
1:ArgType1	-0.100746	0.307400	-0.3277	0.743111	
1:ArgTP1	0.016624	0.074945	0.2218	0.824453	
1:ArgType1:ArgTP1	-0.068656	0.072256	-0.9502	0.342021	
2:(Intercept)	-8.995567	0.203890	-44.1198	< 2.2e-16	***
2:ArgAnim1	-0.379069	0.363861	-1.0418	0.297507	
2:ArgRlz1	-0.300508	0.498766	-0.6025	0.546840	
2:ArgRlz2	-0.134489	0.317900	-0.4231	0.672256	
2:ArgType1	-8.537989	0.203890	-41.8755	< 2.2e-16	***
2:ArgTP1	-0.309710	0.067834	-4.5657	4.978e-06	***
2:ArgType1:ArgTP1	-0.113836	0.067834	-1.6782	0.093316	.
3:(Intercept)	-10.221565	0.247735	-41.2600	< 2.2e-16	***
3:ArgAnim1	8.938296	0.247736	36.0800	< 2.2e-16	***
3:ArgRlz1	-1.193408	0.561418	-2.1257	0.033528	*
3:ArgRlz2	-0.173807	0.462941	-0.3754	0.707333	
3:ArgType1	-1.148393	0.450811	-2.5474	0.010853	*
3:ArgTP1	-0.055612	0.099780	-0.5573	0.577294	
3:ArgType1:ArgTP1	-0.126191	0.102020	-1.2369	0.216117	

E.3.2. Binary fixed-effects models – without considering the verbs (Sections 3.4.2 & 3.4.3)

E.3.2.1. Zero

```
318 tokens
Distribution:
             1     196
             0     122

Model selection:
Z ~ ArgType * ArgTP1 + ArgAnim + ArgRlz
               Df Deviance    AIC     LRT  Pr(>Chi)
<none>              374.10 388.10
ArgAnim         1   385.26 397.26 11.1621 0.0008348 ***
ArgRlz          2   387.40 397.40 13.3005 0.0012937 **
ArgType:ArgTP1  1   374.10 386.10  0.0002 0.9888247

Z ~ ArgType + ArgAnim + ArgRlz + ArgTP1
        Df Deviance    AIC     LRT  Pr(>Chi)
<none>       374.10 386.10
ArgType  1   404.16 414.16 30.0675 4.173e-08 ***
ArgAnim  1   385.58 395.58 11.4863 0.0007011 ***
ArgRlz   2   387.66 395.66 13.5670 0.0011323 **
ArgTP1   1   375.56 385.56  1.4596 0.2269898

Z ~ ArgType + ArgAnim + ArgRlz
        Df Deviance    AIC    LRT  Pr(>Chi)
<none>       375.56 385.56
ArgType  1   409.27 417.27 33.717 6.375e-09 ***
ArgAnim  1   385.58 393.58 10.029  0.001541 **
ArgRlz   2   388.32 394.32 12.765  0.001691 **

Summary of results:
Call:
glm(formula = Z ~ ArgType + ArgAnim + ArgRlz, family = "binomial", data
= AltDS)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-2.1042  -1.2564   0.5477   0.9732   1.5941

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)   0.7039     0.1509   4.664 3.11e-06 ***
ArgType1      0.9572     0.1813   5.279 1.30e-07 ***
ArgAnim1     -0.5625     0.1879  -2.993  0.00276 **
ArgRlz1       0.7228     0.2517   2.871  0.00409 **
ArgRlz2      -0.1254     0.1974  -0.635  0.52550
---
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 423.46  on 317  degrees of freedom
Residual deviance: 375.56  on 313  degrees of freedom
AIC: 385.56
```

E.3.2.2. /e/-series

```
318 tokens
Distribution:
               1   53
               0  265

Model selection:
GEN ~ ArgAnim + ArgRlz + ArgType * ArgTP1
               Df Deviance    AIC    LRT Pr(>Chi)
<none>              266.59 280.59
ArgAnim         1   272.76 284.76 6.1782  0.01293 *
ArgRlz          2   271.71 281.71 5.1199  0.07731 .
ArgType:ArgTP1  1   267.40 279.40 0.8125  0.36737

GEN ~ ArgType + ArgAnim + ArgRlz + ArgTP1
        Df Deviance    AIC    LRT Pr(>Chi)
<none>       267.40 279.40
ArgType  1   267.91 277.91 0.5099 0.475203
ArgAnim  1   274.70 284.70 7.3052 0.006876 **
ArgRlz   2   272.35 280.35 4.9529 0.084043 .
ArgTP1   1   267.40 277.40 0.0006 0.980812

GEN ~ ArgAnim + ArgRlz
        Df Deviance    AIC     LRT  Pr(>Chi)
<none>       267.94 275.94
ArgAnim  1   283.82 289.82 15.8786 6.754e-05 ***
ArgRlz   2   272.69 276.69  4.7539   0.09283 .

Summary of results:
Call:
glm(formula = GEN ~ ArgAnim + ArgRlz, data = AltDS)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-0.8159  -0.7684  -0.4185  -0.3913   2.2838

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  -2.0307     0.2138  -9.496  < 2e-16 ***
ArgAnim1      0.7312     0.1990   3.675 0.000238 ***
ArgRlz1      -0.6012     0.2966  -2.027 0.042680 *
ArgRlz2       0.3705     0.2399   1.545 0.122446
---
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 286.56  on 317  degrees of freedom
Residual deviance: 267.94  on 314  degrees of freedom
AIC: 275.94
```

E.3.2.3. /o/-series

```
318 tokens
Distribution:
                   1   46
                   0  272
```

Model selection:
NB. ArgType is not relevant here, as the data only include P arguments.

```
LOC ~ ArgAnim * ArgRlz + ArgAnim * ArgTP1
               Df Deviance    AIC     LRT Pr(>Chi)
<none>              185.12 201.12
ArgAnim:ArgRlz  2   188.18 200.18 3.05702   0.2169
ArgAnim:ArgTP1  1   185.68 199.68 0.56181   0.4535

LOC ~ ArgRlz + ArgAnim * ArgTP1
               Df Deviance    AIC     LRT Pr(>Chi)
<none>              188.18 200.18
ArgRlz          2   190.42 198.42 2.24341   0.3257
ArgAnim:ArgTP1  1   188.83 198.83 0.65218   0.4193

LOC ~ ArgAnim + ArgRlz + ArgTP1
        Df Deviance    AIC     LRT  Pr(>Chi)
<none>       188.83 198.83
ArgAnim  1   217.60 225.60 28.7708 8.147e-08 ***
ArgRlz   2   190.91 196.91  2.0802   0.35341
ArgTP1   1   193.70 201.70  4.8661   0.02739 *

LOC ~ ArgRlz
       Df Deviance    AIC    LRT Pr(>Chi)
<none>      253.49 259.49
ArgRlz  2   262.87 264.87 9.3801 0.009186 **

LOC ~ ArgRlz + ArgTP1
       Df Deviance    AIC    LRT  Pr(>Chi)
<none>      217.60 225.60
ArgRlz  2   223.26 227.26  5.656   0.05912 .
ArgTP1  1   253.49 259.49 35.891 2.087e-09 ***
```

Summary of results:

```
Call:
glm(formula = LOC ~ ArgAnim + ArgTP1,family = "binomial",data = AltDS)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-0.9827  -0.5769  -0.1668  -0.1055   2.9262

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  -1.8794     0.3661  -5.134 2.84e-07 ***
ArgAnim1     -1.4025     0.3240  -4.328 1.50e-05 ***
ArgTP1       -0.2464     0.1199  -2.054   0.0399 *
---
Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 262.87  on 317  degrees of freedom
Residual deviance: 190.91  on 315  degrees of freedom
AIC: 196.9
```

➔ Removing animacy and including argument realization

```
Call:
glm(formula = LOC ~ ArgRlz + ArgTP1,family = "binomial", data = AltDS)
```

```
Deviance Residuals:
    Min       1Q   Median       3Q      Max
-0.8978  -0.7083  -0.3115  -0.1357   2.3612

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  -1.1625     0.2601  -4.470 7.83e-06 ***
ArgRlz1      -0.8594     0.4209  -2.042   0.0411 *
ArgRlz2       0.4620     0.2709   1.706   0.0880 .
ArgTP1       -0.4898     0.1069  -4.580 4.65e-06 ***
---
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 262.87  on 317  degrees of freedom
Residual deviance: 217.60  on 314  degrees of freedom
AIC: 225.6
```

E.3.2.4. /a/-series

318 tokens
Distribution:

```
                   1   23
                   0  295
```

Model selection:

```
PAT ~ ArgAnim + ArgRlz + ArgType * ArgTP1
               Df Deviance    AIC    LRT  Pr(>Chi)
<none>              100.30 114.30
ArgAnim         1   139.23 151.23 38.935 4.381e-10 ***
ArgRlz          2   111.06 121.06 10.769  0.004586 **
ArgType:ArgTP1  1   101.66 113.66  1.367  0.242268

PAT ~ ArgRlz + ArgType * ArgTP1
               Df Deviance    AIC     LRT  Pr(>Chi)
<none>              139.23 151.23
ArgRlz          2   154.02 162.02 14.7879 0.0006149 ***
ArgType:ArgTP1  1   145.77 155.77  6.5351 0.0105769 *

Model 1: PAT ~ ArgAnim + ArgRlz + ArgType * ArgTP1
Model 2: PAT ~ ArgRlz + ArgType * ArgTP1
  Resid. Df Resid. Dev Df Deviance  Pr(>Chi)
1       311     100.30
2       312     139.23 -1  -38.935 4.381e-10 ***

Model 1: PAT ~ ArgTP1 * ArgType + Overt
Model 2: PAT ~ ArgRlz + ArgType * ArgTP1
  Resid. Df Resid. Dev Df Deviance Pr(>Chi)
1       313     139.27
2       312     139.23  1 0.037213    0.847

PAT ~ ArgTP1 * ArgType + Overt
               Df Deviance    AIC     LRT  Pr(>Chi)
<none>              139.27 149.27
Overt           1   154.02 162.02 14.7507 0.0001227 ***
ArgTP1:ArgType  1   145.88 153.88  6.6145 0.0101153 *
```

Summary of results:

```
Call:
glm(formula = PAT ~ ArgRlz + ArgType * ArgTP1, family = "binomial", data
= AltDS)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-1.1382  -0.4074  -0.2231  -0.1947   2.9566

Coefficients:
                Estimate Std. Error z value Pr(>|z|)
(Intercept)     -3.10742    0.43583  -7.130    1e-12 ***
ArgRlz1         -0.52030    0.54018  -0.963   0.3355
ArgRlz2         -0.69539    0.42485  -1.637   0.1017
ArgType1        -0.05350    0.37967  -0.141   0.8879
ArgTP1           0.12435    0.08503   1.462   0.1436
ArgType1:ArgTP1 -0.22472    0.08827  -2.546   0.0109 *
---
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 165.12  on 317  degrees of freedom
Residual deviance: 139.23  on 312  degrees of freedom
AIC: 151.23
```

- ➔ With animacy included:

```
Call:
glm(formula = PAT ~ ArgAnim + ArgRlz + ArgType * ArgTP1, family =
"binomial", data = AltDS)

Deviance Residuals:
     Min        1Q    Median        3Q       Max
-1.49882  -0.32946  -0.11467  -0.00004   3.01394

Coefficients:
                 Estimate Std. Error z value Pr(>|z|)
(Intercept)     -11.72125  728.34793  -0.016   0.9872
ArgAnim1          9.95361  728.34785   0.014   0.9891
ArgRlz1          -0.96356    0.54895  -1.755   0.0792 .
ArgRlz2          -0.24204    0.45390  -0.533   0.5939
ArgType1         -0.98320    0.42878  -2.293   0.0218 *
ArgTP1           -0.05081    0.09428  -0.539   0.5899
ArgType1:ArgTP1  -0.11255    0.09677  -1.163   0.2448
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 165.12  on 317  degrees of freedom
Residual deviance: 100.30  on 311  degrees of freedom
AIC: 114.3
```

- ➔ Replacing ArgRlz by a binary variable: Overt(NP+Pro) = +1 vs. Null = -1

```
Call:
glm(formula = PAT ~ Overt + ArgType * ArgTP1, family = "binomial", data
= AltDS)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-1.1369  -0.4089  -0.2200  -0.1907   2.9378

Coefficients:
                Estimate Std. Error z value Pr(>|z|)
(Intercept)     -1.89016    0.40679  -4.646 3.38e-06 ***
Overt1          -1.86227    0.52305  -3.560  0.00037 ***
ArgType1        -0.04295    0.37555  -0.114  0.90895
ArgTP1           0.12446    0.08509   1.463  0.14359
ArgType1:ArgTP1 -0.22576    0.08811  -2.562  0.01040 *
---
Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 165.12  on 317  degrees of freedom
Residual deviance: 139.27  on 313  degrees of freedom
AIC: 149.27
```

➔ With animacy included:

```
Call:
glm(formula = PAT ~ ArgTP1 * ArgType + Overt + ArgAnim, family =
"binomial", data = AltDS)

Deviance Residuals:
     Min        1Q    Median        3Q       Max
-1.51292  -0.32861  -0.13168  -0.00004   2.86970

Coefficients:
                 Estimate Std. Error z value Pr(>|z|)
(Intercept)     -10.47810  721.93744  -0.015  0.98842
ArgTP1           -0.04691    0.09374  -0.500  0.61682
ArgType1         -0.99751    0.42697  -2.336  0.01948 *
Overt1           -1.73028    0.58635  -2.951  0.00317 **
ArgAnim1          9.90777  721.93737   0.014  0.98905
ArgTP1:ArgType1  -0.11370    0.09640  -1.180  0.23818
---
Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 165.12  on 317  degrees of freedom
Residual deviance: 100.94  on 312  degrees of freedom
AIC: 112.94
```

E.3.3. Binary fixed-effects models – fitted to matching verbs only (Section 3.4.4)

E.3.3.1. Zero

Same dataset as above.

E.3.3.2. /e/-series

```
163 tokens
Distribution:
                   1   53
                   0   110

Model selection:
GEN ~ ArgType + ArgAnim + ArgRlz + ArgTP1
        Df Deviance    AIC    LRT Pr(>Chi)
<none>       192.46 204.46
ArgType  1   199.98 209.98 7.5229 0.006092 **
ArgAnim  1   194.86 204.86 2.4068 0.120812
ArgRlz   2   199.07 207.07 6.6176 0.036560 *
ArgTP1   1   192.46 202.46 0.0083 0.927296

GEN ~ ArgType + ArgRlz
        Df Deviance    AIC    LRT Pr(>Chi)
<none>       195.59 203.59
ArgType  1   200.25 206.25 4.6540  0.03098 *
ArgRlz   2   201.41 205.41 5.8115  0.05471 .

Summary of results:
Call:
glm(formula = GEN ~ ArgType + ArgRlz, family = "binomial", data = kdvg0)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-1.2910  -0.9137  -0.8498   1.1154   1.9561

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  -0.6011     0.2190  -2.745  0.00606 **
ArgType1     -0.4604     0.2128  -2.163  0.03053 *
ArgRlz1       0.2883     0.2508   1.149  0.25041
ArgRlz2      -0.6921     0.3101  -2.232  0.02564 *
---
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 205.61  on 162  degrees of freedom
Residual deviance: 195.60  on 159  degrees of freedom
AIC: 203.6
```

E.3.3.3. /o/-series

```
139 tokens
Distribution:
                   1  46
                   0  93
```

```
Model selection:
LOC ~ ArgAnim * ArgTP1 + ArgRlz
               Df Deviance    AIC     LRT Pr(>Chi)
<none>              132.95 144.95
ArgRlz          2   133.84 141.84 0.88811   0.6414
ArgAnim:ArgTP1  1   133.89 143.89 0.94188   0.3318

LOC ~ ArgAnim + ArgTP1 + ArgRlz
        Df Deviance    AIC     LRT  Pr(>Chi)
<none>       133.89 143.89
ArgAnim  1   145.44 153.44 11.5538 0.0006761 ***
ArgTP1   1   137.41 145.41  3.5199 0.0606343 .
ArgRlz   2   134.81 140.81  0.9204 0.6311663

LOC ~ ArgAnim * ArgTP1
               Df Deviance    AIC     LRT Pr(>Chi)
<none>              133.84 141.84
ArgAnim:ArgTP1  1   134.81 140.81 0.97414   0.3236

LOC ~ ArgAnim + ArgTP1
        Df Deviance    AIC    LRT  Pr(>Chi)
<none>       134.81 140.81
ArgAnim  1   150.26 154.26 15.449 8.475e-05 ***
ArgTP1   1   138.35 142.35  3.540    0.0599 .


Summary of results:
Call:
glm(formula = LOC ~ ArgAnim + ArgTP1, family = "binomial", data = kdvl0)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-1.3017  -0.9107  -0.2786   1.0580   2.4651

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  -0.8706     0.4027  -2.162 0.030618 *
ArgAnim1     -1.1581     0.3453  -3.354 0.000797 ***
ArgTP1       -0.2401     0.1342  -1.789 0.073594 .
---
Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 176.49  on 138  degrees of freedom
Residual deviance: 134.81  on 136  degrees of freedom
AIC: 140.81
```

E.3.3.4. /a/-series

60 tokens
Distribution:

```
               1  23
               0  37
```

```
Model selection:
PAT ~ ArgType * ArgTP1 + ArgRlz + ArgAnim
               Df Deviance    AIC     LRT  Pr(>Chi)
<none>              22.008 36.008
```

```
ArgRlz          2   23.931 33.931  1.9231    0.3823
ArgAnim         1   53.101 65.101 31.0937 2.459e-08 ***
ArgType:ArgTP1  1   22.937 34.937  0.9295    0.3350

PAT ~ ArgType + ArgAnim + ArgTP1 + ArgRlz
        Df Deviance    AIC     LRT  Pr(>Chi)
<none>       22.937 34.937
ArgType  1   23.471 33.471  0.5336    0.4651
ArgAnim  1   53.923 63.923 30.9860 2.599e-08 ***
ArgTP1   1   23.097 33.097  0.1598    0.6894
ArgRlz   2   25.151 33.151  2.2142    0.3305

PAT ~ ArgType * ArgTP1 + ArgAnim + Overt
               Df Deviance    AIC     LRT  Pr(>Chi)
<none>              22.329 34.329
ArgAnim         1   53.392 63.392 31.0633 2.497e-08 ***
Overt           1   23.931 33.931  1.6020    0.2056
ArgType:ArgTP1  1   23.550 33.550  1.2207    0.2692

PAT ~ ArgType * ArgTP1 + Overt
               Df Deviance    AIC     LRT  Pr(>Chi)
<none>              53.392 63.392
Overt           1   72.258 80.258 18.8656 1.403e-05 ***
ArgType:ArgTP1  1   53.969 61.969  0.5766    0.4476

Model 1: PAT ~ Overt
Model 2: PAT ~ Overt + ArgTP1
  Resid. Df Resid. Dev Df Deviance Pr(>Chi)
1        58     62.089
2        57     54.446  1   7.6427   0.0057 **

Model 1: PAT ~ Overt * ArgTP1
Model 2: PAT ~ Overt + ArgTP1
  Resid. Df Resid. Dev Df Deviance Pr(>Chi)
1        56     54.045
2        57     54.446 -1 -0.40078   0.5267
```

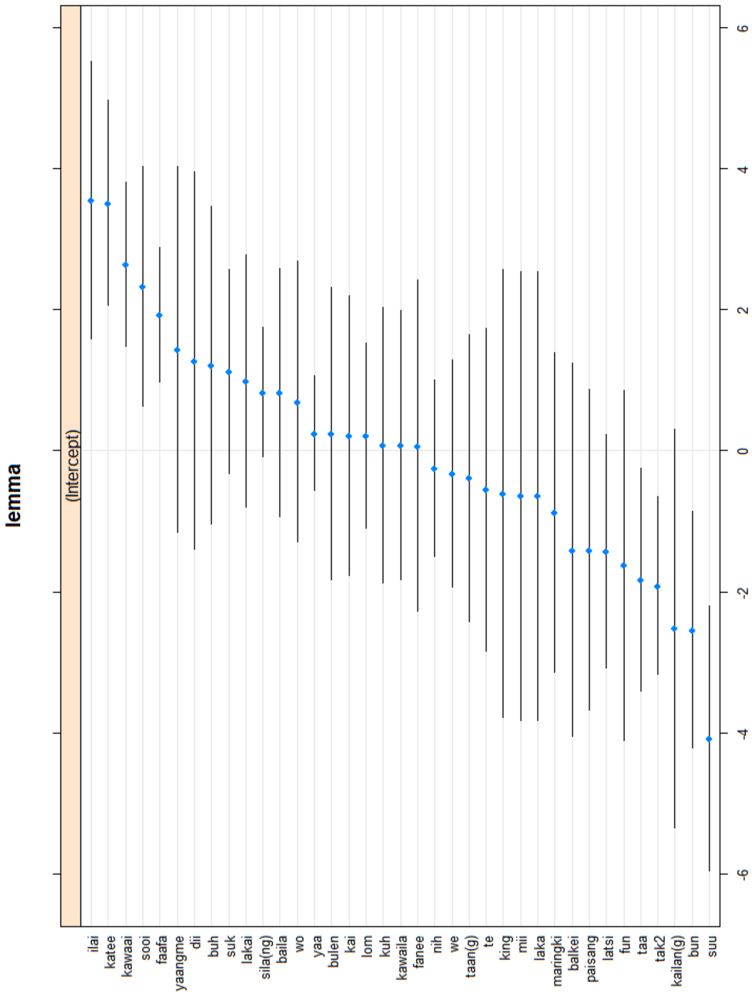
Summary of results:

```
Call:
glm(formula = PAT ~ Overt + ArgTP1, family = "binomial", data = kdvp0)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-1.9917  -0.6049  -0.3433   0.6531   2.0643

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)   0.2409     0.5555   0.434   0.6646
Overt1       -3.0424     0.7795  -3.903  9.5e-05 ***
ArgTP1        0.3986     0.1628   2.449   0.0143 *
---
Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)
    Null deviance: 79.881  on 59  degrees of freedom
Residual deviance: 54.446  on 57  degrees of freedom
AIC: 60.446
```

E.3.4. Binary mixed-effects models – fitted to matching verbs only (Section 3.4.4)

E.3.4.1. Zero

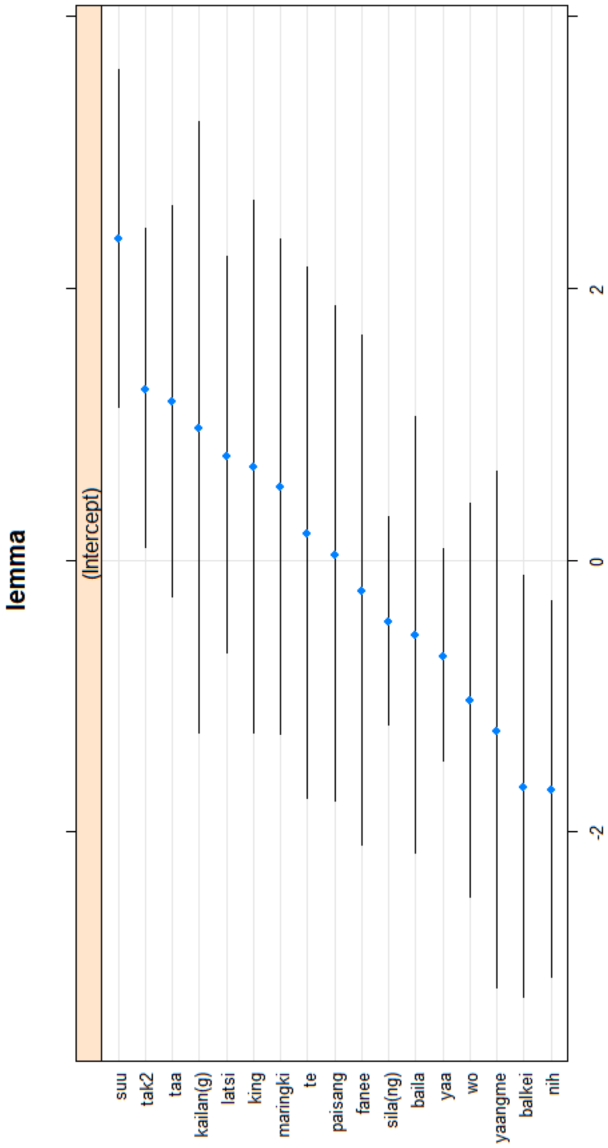
```
Summary of results:
Generalized linear mixed model fit by maximum likelihood (Laplace
Approximation) ['glmerMod']
 Family: binomial  ( logit )
Formula: Z ~ ArgType + ArgAnim + ArgRlz + (1 | lemma)
   Data: AltDS

     AIC      BIC   logLik deviance df.resid
   327.7    350.2   -157.8    315.7      312

Scaled residuals:
    Min      1Q  Median      3Q     Max
-4.4867 -0.5172  0.2249  0.4466  9.1454

Random effects:
 Groups Name        Variance Std.Dev.
 lemma  (Intercept) 4.041    2.01
Number of obs: 318, groups:  lemma, 37

Fixed effects:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  0.39779    0.41304   0.963  0.33550
ArgType1     2.02207    0.35086   5.763 8.25e-09 ***
ArgAnim1    -0.66599    0.25797  -2.582  0.00983 **
ArgRlz1      0.99825    0.34394   2.902  0.00370 **
ArgRlz2     -0.07045    0.26731  -0.264  0.79212
---
Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Correlation of Fixed Effects:
         (Intr) ArgTy1 ArgAn1 ArgRl1
ArgType1  0.034
ArgAnim1 -0.103 -0.497
ArgRlz1   0.256  0.223 -0.225
ArgRlz2  -0.215  0.026  0.258 -0.646
```



lemma
(Intercept)
ilai
katee
kawaai
sooi
faafa
yaangme
dii
buh
suk
lakai
sila(ng)
baila
wo
yaa
bulen
kai
lom
kuh
kawaila
fanee
nih
we
taan(g)
te
king
mii
laka
maringki
balkei
paisang
latsi
fun
taa
tak2
kailan(g)
bun
suu
-6
-4
-2
0
2
4
6

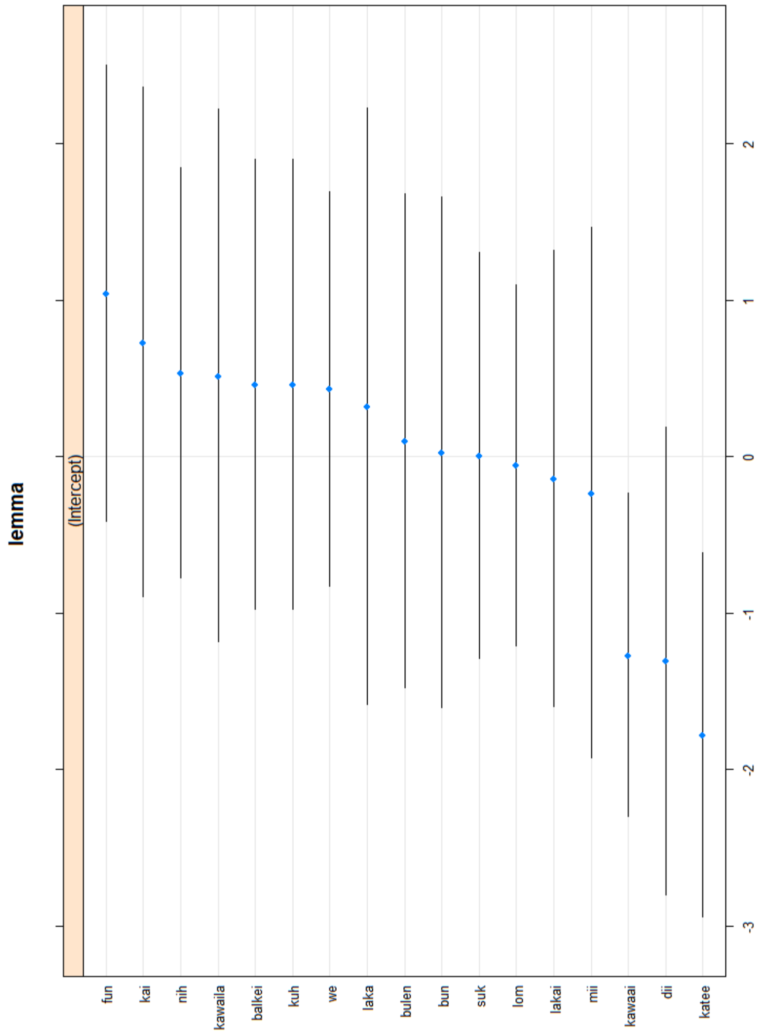
E.3.4.2. /e/-series

```
Generalized linear mixed model fit by maximum likelihood (Laplace
Approximation) ['glmerMod']
 Family: binomial  ( logit )
Formula: GEN ~ ArgType + ArgRlz + (1 | lemma)
   Data: kdvg0

     AIC      BIC   logLik deviance df.resid
   186.6    202.1    -88.3    176.6      158

Scaled residuals:
    Min      1Q  Median      3Q     Max
-1.5274 -0.5725 -0.3501  0.6547  6.9489

Random effects:

 Groups Name        Variance Std.Dev.
 lemma  (Intercept) 1.959    1.4
Number of obs: 163, groups:  lemma, 17

Fixed effects:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  -0.2934     0.4588  -0.640  0.52247
ArgType1     -0.9057     0.3336  -2.715  0.00663 **
ArgRlz1       0.2180     0.3128   0.697  0.48597
ArgRlz2      -0.9981     0.3996  -2.498  0.01249 *
---
Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Correlation of Fixed Effects:
         (Intr) ArgTy1 ArgRl1
ArgType1 -0.302
ArgRlz1  -0.163  0.070
ArgRlz2   0.133  0.252 -0.604
```


lemma
(Intercept)
suu
tak2
taa
kailan(g)
latsi
king
maringki
te
paisang
fanee
sila(ng)
baila
yaa
wo
yaangme
balkei
nih
-2
0
2

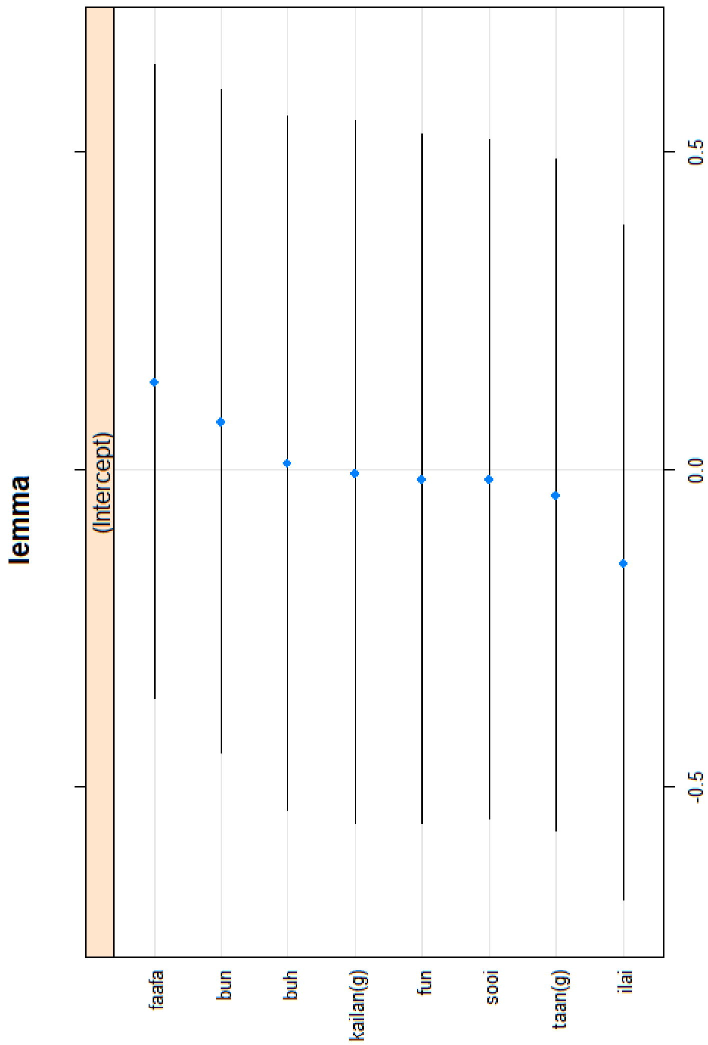
E.3.4.3. /o/-series

```
Generalized linear mixed model fit by maximum likelihood (Laplace
Approximation) ['glmerMod']
 Family: binomial  ( logit )
Formula: LOC ~ ArgAnim + ArgTP1 + (1 | lemma)
   Data: kdvlO

     AIC      BIC   logLik deviance df.resid
   133.0    144.7    -62.5    125.0      135

Scaled residuals:
    Min      1Q  Median      3Q     Max
-1.7328 -0.3953 -0.1760  0.5771  7.5150

Random effects:
 Groups Name        Variance Std.Dev.
 lemma  (Intercept) 1.17     1.082
Number of obs: 139, groups:  lemma, 17

Fixed effects:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  -0.8109     0.5538  -1.464 0.143148
ArgAnim1     -1.4448     0.4137  -3.492 0.000479 ***
ArgTP1       -0.2493     0.1682  -1.482 0.138398
---
Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Correlation of Fixed Effects:
         (Intr) ArgAn1
ArgAnim1  0.648
ArgTP1   -0.589 -0.417
```


lemma
(Intercept)
fun
kai
nih
kawaila
balkei
kuh
we
laka
bulen
bun
suk
lom
lakai
mii
kawaai
dii
katee
-3
-2
-1
0
1
2

E.3.4.4. /a/-series

```
Generalized linear mixed model fit by maximum likelihood (Laplace
Approximation) ['glmerMod']
 Family: binomial  ( logit )
Formula: PAT ~ Overt + ArgTP1 + (1 | lemma)
   Data: kdvp0

     AIC      BIC   logLik deviance df.resid
    62.4     70.8    -27.2     54.4       56

Scaled residuals:
    Min      1Q  Median      3Q     Max
-2.4146 -0.4228 -0.2504  0.5000  2.5958

Random effects:
 Groups Name        Variance Std.Dev.
 lemma  (Intercept) 0.08003  0.2829
Number of obs: 60, groups:  lemma, 8

Fixed effects:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)   0.2197     0.5841   0.376 0.706874
Overt1       -3.0578     0.7943  -3.850 0.000118 ***
ArgTP1        0.3959     0.1643   2.409 0.015995 *
---
Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Correlation of Fixed Effects:
       (Intr) Overt1
Overt1 -0.347
ArgTP1 -0.387 -0.471
```


lemma
(Intercept)
faafa
bun
buh
kailan(g)
fun
sooi
taan(g)
ilai
-0.5
0.0
0.5

E.3.5. Goodness-of-fit of binomial models

E.3.5.1. Zero

R2 values:

```
FE Model                      ME Model
r.squaredLR(BM1_Z)            r.squaredGLMM(BM2_Z)
[1] 0.1398526                              R2m       R2c
attr(,"adj.r.squared")        theoretical 0.3216779 0.6955911
[1] 0.1900282                 delta       0.3016059 0.65218
```

Null Model:

```
BM0_Z=glm(Z~1, family="binomial" , data=AltDS)
```

Observed/predicted contingency tables:

```
Null Model
      1
  0 122
  1 196

FE Model
      0   1
  0  41  81
  1  24 172

ME Model
      0   1
  0  99  23
  1  19 177
```

E.3.5.2. /e/-series

R2 values:

```
FE Model 1                    FE Model 2
r.squaredLR(BM1_GEN)          r.squaredLR(BM2_GEN)
[1] 0.05685856                [1] 0.05956831
attr(,"adj.r.squared")        attr(,"adj.r.squared")
[1] 0.09573979                [1] 0.083110

ME Model
r.squaredGLMM(BM3_GEN)
                  R2m       R2c
theoretical 0.1485881 0.4663572
delta       0.1296784 0.4070072
```

Null Models:

```
BM0_GEN=glm(GEN~1, family="binomial" , data=AltDS)
BM02_GEN=glm(GEN~1, family="binomial" , data=kdvg0)
```

Observed/predicted contingency tables:

Null Model 1
```
     0
  0 265
  1  53
```

FE Model 1
```
     0
  0 265
  1  53
```

Null Model 2
```
     0
  0 110
  1  53
```

FE1 Mode 2
```
    0   1
  0 100  10
  1  40  13
```

ME Model
```
    0   1
  0 99   11
  1 18   35
```

E.3.5.3. /o/-series

R2 Values:

FE Model 1
```
r.squaredLR(BM1_LOC)
[1] 0.202516
attr(,"adj.r.squared")
[1] 0.3600376
```

FE Model 2
```
r.squaredLR(BM2_LOC)
[1] 0.2590459
attr(,"adj.r.squared")
[1] 0.3602474
```

ME Model
```
r.squaredGLMM(BM3_LOC)
                  R2m       R2c
theoretical 0.4508898 0.5949573
delta       0.4079808 0.5383381
```

Null Models:
```
BM0_LOC=glm(LOC~1, family="binomial" , data=AltDS)
BM02_LOC=glm(LOC~1, family="binomial" , data=kdvl0)
```

Observed/predicted contingency tables:

Null Model 1
```
     0
  0 272
  1  46
```

FE Model 1
```
     0
  0 272
  1  46
```

Null Model 2
```
    0
  0 93
  1 46
```

FE1 Mode 2
```
    0  1
  0 61 32
  1 11 3
```

ME Model
```
    0  1
  0 81 12
  1  9 37
```

E.3.5.4. /a/-series

R2 Values:

FE Model 1
```
r.squaredLR(BM1_PAT)
[1] 0.07806917
attr(,"adj.r.squared")
[1] 0.1927519
```

FE Model 2
```
r.squaredLR(BM2_PAT)
[1] 0.3455202
attr(,"adj.r.squared")
[1] 0.4695346
```

ME Model
```
r.squaredGLMM(BM3_PAT)
                  R2m       R2c
theoretical 0.4716706 0.4842174
delta       0.4054077 0.4161919
```

Null Models:
```
BM0_PAT=glm(PAT~1, family="binomial" , data=AltDS)
BM02_PAT=glm(PAT~1, family="binomial" , data=kdvp0)
```

Observed/predicted contingency tables:

Null Model 1

	0
0	295
1	23

FE Model 1

	0	1
0	295	0
1	22	1

Null Model 2

	0
0	37
1	23

FE1 Mode 2

	0	1
0	31	6
1	6	17

ME Model

	0	1
0	30	7
1		