

Supplementary materials for:

On task effects in NLG corpus elicitation: a replication study using mixed effects modeling

Emiel van Miltenburg, Merel van de Kerkhof, Ruud Koolen, Martijn Goudbeek, Emiel Kraemer
Tilburg University

INLG 2019

1 Models

This section provides our R code with the model specifications.

1.1 Requirements

Our code uses the following packages:

- `lme4`, see: Bates et al. 2015
- `lmerTest`, see: Kuznetsova et al. 2017

1.2 Convergent models

Below is the code for the convergent models.

```
1 # Default models
2
3 length.model = lmer(length ~ modality + (1|participant) + (1|image),
4                   data=modality_data)
5
6 pid.model = lmer(PID ~ modality + (1|participant) + (1|image),
7                data=modality_data)
8
9 chars.model = lmer(chars ~ modality + (1|participant) + (1|image),
10                 data=modality_data)
11
12 # Count models - using the poisson distribution
13
14 adverbs.model = glmer(adverbs ~ modality + (1|participant) + (1|image),
15                      data=modality_data, family = "poisson")
16
17 attributives.model = glmer(attributives ~ modality + (1|participant) + (1|image),
18                           data=modality_data, family = "poisson")
19
20 prepositions.model = glmer(prepositions ~ modality + (1|participant) + (1|image),
21                          data=modality_data, family="poisson")
22
23 cop.model = glmer(consciousness_of_projection ~ modality + (1|participant) + (1|image),
24                 data=modality_data, family = "poisson")
25
26 negations.model = glmer(negations ~ modality + (1|participant) + (1|image),
27                       data=modality_data, family = "poisson")
28
29 pq.model = glmer(pseudo_quantifiers ~ modality + (1|participant) + (1|image),
30                data=modality_data, family = "poisson")
```

1.3 Fixing inconvergent models

Some of our models initially did not converge. This section shows how we adapted the models to (hopefully) obtain a stable model.

1.3.1 Number of syllables

The model initially did not converge. Changing the optimizer helped us reach a stable model.

```
1 # Did not converge: with the default optimizer:
2 syll.model = lmer(syllables ~ modality + (1|participant) + (1|image),
3                 data=modality_data)
4
5 # Did converge with bobyqa.
6 syll.model = lmer(syllables ~ modality + (1|participant) + (1|image),
7                 data=modality_data, control=lmerControl(optimizer = "bobyqa"))
```

1.3.2 Self-reference terms

The model for self-reference terms initially did not converge, presumably because of the distribution of the data (many zeroes, some ones, few higher numbers). Using a binomial distribution helped with the sparsity of the data.

```
1 # Does not converge:
2 self_reference.model = glmer(self_reference_words ~ modality + (1|participant) + (1|image),
3                             data=modality_data, family = "poisson")
4
5 # Manipulate data: replace values higher than 1 with 1.
6 modality_data$selfref_capped <- replace(modality_data$self_reference_words,
7                                       modality_data$self_reference_words >= 1,
8                                       1)
9
10 # Does converge
11 selfref_capped.model = glmer(selfref_capped ~ modality + (1|participant) + (1|image),
12                             data=modality_data, family = "binomial")
```

1.3.3 Positive allness terms

The same strategy did not work for positive allness terms.

```
1 # Does not converge:
2 allness.model = glmer(positive_allness ~ modality + (1|participant) + (1|image),
3                       data=modality_data, family = "poisson")
4
5 # Manipulate data: replace values higher than 1 with 1.
6 modality_data$allness_capped <- replace(modality_data$positive_allness,
7                                       modality_data$positive_allness >= 1,
8                                       1)
9
10 # Still does not converge
11 allness.model = glmer(allness_capped ~ modality + (1|participant) + (1|image),
12                       data=modality_data, family = "binomial")
```

2 Results

We provide all the output from the summary function in R, except for the model for allness terms, which did not converge.

2.1 Description length

Below is the output for description length.

```
Linear mixed model fit by REML. t-tests use Satterthwaite's method [
lmerModLmerTest]
Formula: length ~ modality + (1 | participant) + (1 | image)
Data: modality_data
```

REML criterion at convergence: 42838.5

```
Scaled residuals:
  Min      1Q  Median      3Q      Max
-3.8198 -0.5956 -0.0802  0.4716  8.7392
```

```
Random effects:
 Groups      Name      Variance Std.Dev.
 image      (Intercept)  2.527   1.590
 participant (Intercept) 22.591   4.753
 Residual                    22.712   4.766
Number of obs: 7056, groups: image, 307; participant, 93
```

```
Fixed effects:
              Estimate Std. Error    df t value Pr(>|t|)
(Intercept)    12.6250    0.7178 92.6215  17.589 < 2e-16 ***
modalitywritten  2.6304    0.9934 90.5499   2.648  0.00956 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
Correlation of Fixed Effects:
              (Intr)
modltywrtnn -0.711
```

2.2 Adverbs

Below is the output for adverbs.

```
Generalized linear mixed model fit by maximum likelihood (Laplace Approximation)
[glmerMod]
Family: poisson ( log )
Formula: adverbs ~ modality + (1 | participant) + (1 | image)
Data: modality_data
```

```
      AIC      BIC    logLik deviance df.resid
14869.8 14897.2 -7430.9 14861.8     7052
```

```
Scaled residuals:
  Min      1Q  Median      3Q      Max
-1.6834 -0.7229 -0.4784  0.5448  6.9552
```

```
Random effects:
 Groups      Name      Variance Std.Dev.
 image      (Intercept)  0.09163  0.3027
 participant (Intercept) 0.34625  0.5884
Number of obs: 7056, groups: image, 307; participant, 93
```

```
Fixed effects:
```

```

                Estimate Std. Error z value Pr(>|z|)
(Intercept)    -0.63204    0.09197  -6.872 6.33e-12 ***
modalitywritten 0.09211    0.12690   0.726   0.468
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

```

Correlation of Fixed Effects:
      (Intr)
modltywrtn -0.695

```

2.3 Attributive adjectives

Below is the output for attributive adjectives.

```

Generalized linear mixed model fit by maximum likelihood (Laplace Approximation)
 [glmerMod]
Family: poisson ( log )
Formula: attributives ~ modality + (1 | participant) + (1 | image)
Data: modality_data

```

```

      AIC      BIC    logLik deviance df.resid
12334.0 12361.4 -6163.0 12326.0    7052

```

```

Scaled residuals:
  Min       1Q   Median       3Q      Max
-1.6871 -0.5945 -0.4225  0.4572  6.6777

```

```

Random effects:
 Groups      Name      Variance Std.Dev.
 image      (Intercept) 0.4225  0.650
 participant (Intercept) 0.2256  0.475
Number of obs: 7056, groups: image, 307; participant, 93

```

```

Fixed effects:
                Estimate Std. Error z value Pr(>|z|)
(Intercept)    -1.02043    0.08404 -12.143 <2e-16 ***
modalitywritten 0.15068    0.10508   1.434   0.152
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

```

Correlation of Fixed Effects:
      (Intr)
modltywrtn -0.626

```

2.4 Token length (characters)

Below is the output for token length, in terms of characters.

```

Linear mixed model fit by REML. t-tests use Satterthwaite's method [
lmerModLmerTest]
Formula: chars ~ modality + (1 | participant) + (1 | image)
Data: modality_data

```

```

REML criterion at convergence: 14944.9

```

```

Scaled residuals:
  Min       1Q   Median       3Q      Max
-3.1721 -0.6163 -0.1152  0.4483  8.7849

```

```

Random effects:
 Groups      Name      Variance Std.Dev.
 image      (Intercept) 0.13215  0.3635
 participant (Intercept) 0.04207  0.2051

```

```

Residual                0.43246  0.6576
Number of obs: 7056, groups:  image, 307; participant, 93

Fixed effects:
      Estimate Std. Error      df t value Pr(>|t|)
(Intercept)   4.678e+00  3.821e-02  1.473e+02  122.454  <2e-16 ***
modalitywritten 5.047e-03  4.563e-02  8.423e+01   0.111   0.912
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Correlation of Fixed Effects:
      (Intr)
modltywrtn -0.590

```

2.5 Token length (syllables)

Below is the output for token length, measured in syllables.

```

Linear mixed model fit by REML. t-tests use Satterthwaite's method [
lmerModLmerTest]
Formula: syllables ~ modality + (1 | participant) + (1 | image)
Data: modality_data
Control: lmerControl(optimizer = "bobyqa")

REML criterion at convergence: -645.9

```

```

Scaled residuals:
      Min       1Q   Median       3Q      Max
-2.4397 -0.6247 -0.1194  0.4642 10.6550

```

```

Random effects:
Groups      Name      Variance Std.Dev.
image      (Intercept) 0.014390 0.11996
participant (Intercept) 0.003958 0.06292
Residual              0.047530 0.21801
Number of obs: 7056, groups:  image, 307; participant, 93

```

```

Fixed effects:
      Estimate Std. Error      df t value Pr(>|t|)
(Intercept)   1.51933    0.01205  152.61066  126.081  <2e-16 ***
modalitywritten 0.00123    0.01415  82.64442   0.087   0.931
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Correlation of Fixed Effects:
      (Intr)
modltywrtn -0.577

```

2.6 Consciousness-of-projection terms

Below is the output for consciousness-of-projection terms.

```

Generalized linear mixed model fit by maximum likelihood (Laplace Approximation)
 [glmerMod]
Family: poisson ( log )
Formula: consciousness_of_projection ~ modality + (1 | participant) +
(1 | image)
Data: modality_data

      AIC      BIC    logLik deviance df.resid
1445.7  1473.2   -718.9  1437.7    7052

```

```

Scaled residuals:

```

```

      Min      1Q  Median      3Q      Max
-0.6266 -0.1332 -0.0881 -0.0638  9.4834

```

Random effects:

```

Groups      Name          Variance Std.Dev.
image      (Intercept) 0.5035  0.7095
participant (Intercept) 1.5169  1.2316
Number of obs: 7056, groups: image, 307; participant, 93

```

Fixed effects:

```

              Estimate Std. Error z value Pr(>|z|)
(Intercept)   -4.5084    0.2601 -17.332 <2e-16 ***
modalitywritten -0.8523    0.3644  -2.339  0.0193 *
---

```

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Correlation of Fixed Effects:

```

      (Intr)
modltywrtn -0.490

```

2.7 Negations

Below is the output for the use of negations.

Generalized linear mixed model fit by maximum likelihood (Laplace Approximation)

```

[glmerMod]
Family: poisson ( log )
Formula: negations ~ modality + (1 | participant) + (1 | image)
Data: modality_data

```

```

      AIC      BIC    logLik deviance df.resid
876.3    903.7   -434.1    868.3     7052

```

Scaled residuals:

```

      Min      1Q  Median      3Q      Max
-0.4734 -0.0918 -0.0714 -0.0696  9.7975

```

Random effects:

```

Groups      Name          Variance Std.Dev.
image      (Intercept) 0.9206  0.9595
participant (Intercept) 0.6360  0.7975
Number of obs: 7056, groups: image, 307; participant, 93

```

Fixed effects:

```

              Estimate Std. Error z value Pr(>|z|)
(Intercept)   -5.3780    0.2842 -18.92 <2e-16 ***
modalitywritten 0.4376    0.2879  1.52  0.128
---

```

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Correlation of Fixed Effects:

```

      (Intr)
modltywrtn -0.497

```

2.8 Propositional Idea Density

Below is the output for Propositional Idea Density (PID).

Linear mixed model fit by REML. t-tests use Satterthwaite's method [

```

lmerModLmerTest]
Formula: PID ~ modality + (1 | participant) + (1 | image)
Data: modality_data

```

REML criterion at convergence: -11805.5

Scaled residuals:

Min	1Q	Median	3Q	Max
-4.7320	-0.6034	0.0159	0.6176	5.6100

Random effects:

Groups	Name	Variance	Std.Dev.
image	(Intercept)	0.001626	0.04032
participant	(Intercept)	0.000807	0.02841
Residual		0.009995	0.09998

Number of obs: 7056, groups: image, 307; participant, 93

Fixed effects:

	Estimate	Std. Error	df	t value	Pr(> t)
(Intercept)	4.434e-01	5.041e-03	1.262e+02	87.959	<2e-16 ***
modalitywritten	2.350e-03	6.403e-03	9.038e+01	0.367	0.714

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Correlation of Fixed Effects:

(Intr)
modltywrtn -0.623

2.9 Pseudo-quantifiers

Below is the output for pseudo-quantifiers.

Generalized linear mixed model fit by maximum likelihood (Laplace Approximation) [glmerMod]

Family: poisson (log)

Formula: pseudo_quantifiers ~ modality + (1 | participant) + (1 | image)

Data: modality_data

AIC	BIC	logLik	deviance	df.resid
2714.3	2741.7	-1353.1	2706.3	7052

Scaled residuals:

Min	1Q	Median	3Q	Max
-1.1014	-0.2075	-0.1351	-0.0938	8.2960

Random effects:

Groups	Name	Variance	Std.Dev.
image	(Intercept)	1.755	1.3246
participant	(Intercept)	0.611	0.7816

Number of obs: 7056, groups: image, 307; participant, 93

Fixed effects:

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	-4.1827	0.1907	-21.929	<2e-16 ***
modalitywritten	0.4589	0.2006	2.288	0.0222 *

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Correlation of Fixed Effects:

(Intr)
modltywrtn -0.529

2.10 Self-reference terms

Below is the output for the use of self-reference terms.

Generalized linear mixed model fit by maximum likelihood (Laplace Approximation)

```
[glmerMod]
Family: binomial ( logit )
Formula: selfref_capped ~ modality + (1 | participant) + (1 | image)
Data: modality_data
```

AIC	BIC	logLik	deviance	df.resid
799.3	826.7	-395.6	791.3	7052

```
Scaled residuals:
  Min      1Q  Median      3Q      Max
-3.0981 -0.0920 -0.0235 -0.0109 10.4749
```

```
Random effects:
 Groups      Name          Variance Std.Dev.
 image      (Intercept)  0.1653  0.4066
 participant (Intercept) 14.4782  3.8050
Number of obs: 7056, groups: image, 307; participant, 93
```

```
Fixed effects:
              Estimate Std. Error z value Pr(>|z|)
(Intercept)   -6.6485    0.8539  -7.786 6.93e-15 ***
modalitywritten -2.2905    1.0100  -2.268  0.0233 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
Correlation of Fixed Effects:
              (Intr)
modltywrtnn -0.412
```

2.11 Prepositions

Below is the output for the use of prepositions.

```
Generalized linear mixed model fit by maximum likelihood (Laplace Approximation)
[glmerMod]
Family: poisson ( log )
Formula: prepositions ~ modality + (1 | participant) + (1 | image)
Data: modality_data
```

AIC	BIC	logLik	deviance	df.resid
21611.2	21638.7	-10801.6	21603.2	7052

```
Scaled residuals:
  Min      1Q  Median      3Q      Max
-1.8047 -0.4847 -0.0875  0.4117  3.7791
```

```
Random effects:
 Groups      Name          Variance Std.Dev.
 image      (Intercept)  0.03285  0.1812
 participant (Intercept) 0.10342  0.3216
Number of obs: 7056, groups: image, 307; participant, 93
```

```
Fixed effects:
              Estimate Std. Error z value Pr(>|z|)
(Intercept)    0.52614    0.05039  10.441 < 2e-16 ***
modalitywritten 0.26030    0.06908   3.768 0.000165 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
Correlation of Fixed Effects:
              (Intr)
modltywrtnn -0.698
```


References

- D. Bates, M. Mächler, B. Bolker, and S. Walker. Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67(1):1–48, 2015. doi: 10.18637/jss.v067.i01.
- A. Kuznetsova, P. B. Brockhoff, and R. H. B. Christensen. lmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software*, 82(13):1–26, 2017. doi: 10.18637/jss.v082.i13.