**Interpreting mixed effects model outputs**

In this paper we use linear mixed effects models to analyse eye-tracking data. We chose this method of analysis as it provides a way to overcome the “language as fixed effect fallacy” (Clark, 1973) by allowing us to include both participants and stimulus items as random effects. Using regression modelling also allows us to treat lexical frequency, idiom frequency and idiom subjective familiarity like the continuous variables that they are, instead of categorising them into arbitrary bins (e.g. high vs. low). The data is reported in a way that will allow readers to see the structure of each model by presenting a full list of all of the fixed effects (e.g. Participant group, natives vs. non-natives; Phrase type, idioms vs. controls; category of idioms, English vs. Swedish vs. congruent) and covariates considered (e.g. word length, word frequency), as well as a full list of the random effects structure (random variance attributable to each participant, and to each item).

For fixed effects and covariates we report the following:

The coefficient (β), which is the numerical size of the effect for any given fixed effect/covariate. This demonstrates what difference the fixed effect has compared to the Intercept value (the baseline value for the dataset). For example, in a lexical decision task, if a model of native vs. non-native speakers has an intercept value of 500ms (mean time for native speakers to make lexical decisions), a coefficient of 100ms for the effect of ‘group’ would mean that non-native speakers on average take 100ms longer to respond. For continuous predictors, the coefficient indicates the change caused by an increase of one unit (e.g. one letter for word length). For this reason predictors on large scales, such as frequency, are log transformed, both to reduce skewing and to make predictors comparable within the model.

Standard error (SE), which is calculated as the standard deviation of the dataset divided by the square root of the sample size.

The *t*-value (*t*) is comparable to values such as *z*-scores, *f*-scores or chi-squared values in other statistical analyses. From this we can infer a value of significance (which in most statistical methods is reported as the *p*-value). Generally a *t*-value of 2 (or -2) is taken to be the threshold of significance. In our results we also use the lmerTest package in R (version 2.0-11; Kuznetsova, Brockhoff & Christensen, 2014) to estimate *p*-values, which we report using the convention of: \*\*\* p < .001, \*\* p < .01, \* p< .05. (N.B for models considering binary data or non-normally distributed numerical data, as used for the analysis of skipping rates and fixation counts, generalised linear models were used, and these report *z*-values rather than *t­*-values. The calculation and reporting of significance for these is the same, although the interpretation of the coefficients changes.)

Each model therefore reports information on the fixed effects, indicating the magnitude of the effects along with any significant values, and any covariates that were significant. Models were constructed using a stepwise comparison procedure, so an initial model was created including only the fixed effects of interest, then covariates were included one by one and the resulting models compared. Any covariates that significantly improved the fit of the model were retained, and non-significant covariates were dropped.

The final models are reported in full for each of the eye-tracking measures considered in this study. Separate models are presented for phrase-level and word-level analysis, as detailed in the main text.

**Mixed effects model outputs referenced in the main paper**

|  |  |  |  |
| --- | --- | --- | --- |
| Swedish NS | First pass reading time | Total reading time | Fixation count |
| Fixed effects: | β | SE | t | β | SE | t | β | SE | z |
| Intercept | 6.62 | 0.24 | 28.17 | 7.52 | 0.23 | 32.42 | 1.95 | 0.23 | 8.44 |
| Type | -0.04 | 0.05 | -0.83 | -0.10 | 0.04 | -2.65\*\* | -0.08 | 0.04 | -1.96\* |
| Condition: Congruent | -0.05 | 0.06 | -0.83 | -0.15 | 0.06 | -2.37\* | -0.11 | 0.06 | 0.06+ |
| Condition: English | 0.02 | 0.06 | 0.30 | -0.18 | 0.06 | -2.89\*\* | -0.17 | 0.06 | -2.75\*\* |
| Type\*Condition: Congruent | 0.07 | 0.07 | 0.98 | 0.00 | 0.05 | 0.02 | -0.01 | 0.06 | -0.23 |
| Type\*Condition: English | -0.04 | 0.07 | -0.57 | 0.05 | 0.05 | 0.97 | 0.03 | 0.06 | 0.49 |
| Control predictors: |  |  |  |  |  |  |  |  |  |
| Word 1 Length | 0.00 | 0.01 | 0.01 | -0.02 | 0.01 | -1.92\* | -0.02 | 0.01 | -1.21 |
| Word 1 Frequency (log) | -0.04 | 0.01 | -2.95\*\* | -0.02 | 0.01 | -2.50\* | -0.02 | 0.01 | -1.65 |
| Word 2 Length | 0.04 | 0.02 | 2.20\* | 0.02 | 0.02 | 1.23 | 0.02 | 0.02 | 1.18 |
| Word 2 Frequency (log) | -0.02 | 0.02 | -1.35 | -0.03 | 0.02 | -2.03\* | -0.02 | 0.02 | -1.48 |
| Random effects | Variance | Variance | Variance |
| Item | 0.025 |  |  | 0.044 |  |  | 0.034 |  |
| Subject | 0.045 |  |  | 0.108 |  |  | 0.099 |  |
| Subject | Type | 0.001 |  |  | 0.002 |  |  | 0.000 |  |
| Subject | Condition: Congruent | 0.000 |  |  | 0.001 |  |  | 0.003 |  |
| Subject | Condition: English | 0.000 |  |  | 0.006 |  |  | 0.005 |  |
| Residual | 0.327 |  |  | 0.160 |  |  | n/a |  |  |
| English NS | First pass reading time | Total reading time | Fixation count |
| Fixed effects: | β | SE | t | β | SE | t | β | SE | z |
| Intercept | 5.97 | 0.19 | 30.65 | 6.06 | 0.23 | 26.45 | 0.80 | 0.24 | 3.38 |
| Type | -0.16 | 0.04 | -3.78\*\*\* | -0.19 | 0.04 | -4.34\*\*\* | -0.17 | 0.06 | -2.94\*\* |
| Condition: Congruent | -0.05 | 0.05 | -0.99 | -0.04 | 0.06 | -0.75 | -0.02 | 0.06 | -0.37 |
| Condition: Swedish | -0.01 | 0.06 | -0.18 | 0.09 | 0.06 | 1.47 | 0.08 | 0.07 | 1.28 |
| Type\*Condition: Congruent | 0.04 | 0.05 | 0.76 | 0.07 | 0.06 | 1.22 | 0.06 | 0.08 | 0.76 |
| Type\*Condition: Swedish | 0.14 | 0.06 | 2.57\* | 0.41 | 0.06 | 7.22\*\*\* | 0.42 | 0.08 | 5.56\*\*\* |
| Control predictors: |  |  |  |  |  |  |  |  |  |
| Word 1 Length | 0.02 | 0.01 | 1.41 | 0.01 | 0.01 | 0.80 | 0.00 | 0.01 | 0.13 |
| Word 1 Frequency (log) | 0.00 | 0.01 | 0.34 | 0.01 | 0.01 | 1.12 | 0.01 | 0.01 | 0.73 |
| Word 2 Length | 0.02 | 0.01 | 1.24 | 0.01 | 0.02 | 0.81 | 0.02 | 0.02 | 1.19 |
| Word 2 Frequency (log) | -0.02 | 0.01 | -1.74+ | -0.01 | 0.02 | -0.61 | -0.00 | 0.01 | -0.05 |
| Random effects | Variance | Variance | Variance |
| Item | 0.020 |  |  | 0.037 |  |  | 0.020 |  |
| Subject | 0.038 |  |  | 0.074 |  |  | 0.043 |  |
| Subject | Type | 0.005 |  |  | 0.009 |  |  | 0.002 |  |
| Subject | Condition: Congruent | 0.010 |  |  | 0.003 |  |  | 0.003 |  |
| Subject | Condition: Swedish | 0.022 |  |  | 0.006 |  |  | 0.004 |  |
| Residual | 0.180 |  |  | 0.189 |  |  | n/a |  |  |

*Table A.* Separate mixed effects model analysis for Swedish speakers (top) and English native speakers (bottom), phrase level measures. For condition, native language is taken to be the baseline. Table displays coefficients (β), standard error (SE) and t-values (z-values for fixation count), with significance values estimated by the lmerTest package in R (version 2.0-11; Kuznetsova, Brockhoff & Christensen, 2014): \*\*\* p < .001, \*\* p < .01, \* p< .05

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | Likelihood of skipping | First fixation duration | Gaze duration | Total reading time | Regression path duration |
| Fixed effects:  | β | SE | z | β | SE | t | β | SE | t | β | SE | t | β | SE | t |
| Intercept | -3.49 | 1.30 | -2.68 | 5.48 | 0.10 | 54.33 | 5.60 | 0.13 | 41.87 | 6.02 | 0.19 | 31.91 | 6.56 | 0.23 | 28.54 |
| Type: Idiom | 1.72 | 0.58 | 2.96\*\* | -0.01 | 0.03 | -0.34 | -0.01 | 0.04 | -0.32 | -0.11 | 0.05 | -2.27\* | -0.08 | 0.06 | -1.31 |
| Condition: Congruent | 0.83 | 0.68 | 1.22 | -0.07 | 0.04 | -1.97\* | -0.09 | 0.05 | -1.93\* | -0.24 | 0.07 | -3.58\*\*\* | -0.27 | 0.08 | -3.38\*\* |
| Condition: English | 2.29 | 0.61 | 3.72\*\*\* | -0.06 | 0.04 | -1.52 | -0.04 | 0.05 | -0.82 | -0.23 | 0.07 | -3.26\*\* | -0.32 | 0.08 | -4.02\*\*\* |
| Type\*Condition: Congruent | -0.26 | 0.71 | -0.37 | 0.02 | 0.05 | 0.50 | 0.03 | 0.05 | 0.65 | 0.06 | 0.07 | 0.91 | -0.04 | 0.08 | -0.53 |
| Type\*Condition: English | -1.78 | 0.65 | -2.74\*\* | 0.02 | 0.05 | 0.47 | -0.04 | 0.05 | -0.74 | 0.05 | 0.07 | 0.73 | 0.03 | 0.09 | 0.36 |
| Control predictors: |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| Word 2 Length | -0.40 | 0.13 | -3.09 | 0.01 | 0.01 | 1.46 | 0.05 | 0.01 | 3.64\*\*\* | 0.07 | 0.02 | 3.77\*\*\* | 0.04 | 0.02 | 1.94\* |
| Word 2 Frequency (log) | 0.08 | 0.11 | 0.79 | -0.01 | 0.01 | -0.87 | -0.03 | 0.01 | -2.52\* | -0.03 | 0.02 | -2.16\* | -0.04 | 0.02 | -2.00\* |
| Random effects | Variance | Variance | Variance | Variance | Variance |
| Item | 0.853 |  |  | 0.005 |  |  | 0.015 |  |  | 0.032 |  |  | 0.041 |  |  |
| Subject | 0.443 |  |  | 0.018 |  |  | 0.029 |  |  | 0.059 |  |  | 0.154 |  |  |
| Subject | Type | n/a |  |  | 0.000 |  |  | 0.007 |  |  | 0.001 |  |  | 0.007 |  |  |
| Subject | Condition: Congruent | n/a |  |  | 0.002 |  |  | 0.006 |  |  | 0.009 |  |  | 0.012 |  |  |
| Subject | Condition: English | n/a |  |  | 0.002 |  |  | 0.006 |  |  | 0.019 |  |  | 0.014 |  |  |
| Residual | n/a |  |  | 0.113 |  |  | 0.138 |  |  | 0.251 |  |  | 0.396 |  |  |

*Table B.* Separate mixed effects model analysis for Swedish native speakers, word level measures. For condition, native language is taken to be the baseline. Table displays coefficients (β), standard error (SE) and t-values (z-values for likelihood of skipping), with significance values estimated by the lmerTest package in R (version 2.0-11; Kuznetsova, Brockhoff & Christensen, 2014): \*\*\* p < .001, \*\* p < .01, \* p< .05

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | Likelihood of skipping | First fixation duration | Gaze duration | Total reading time | Regression path duration |
| Fixed effects:  | β | SE | z | β | SE | t | β | SE | t | β | SE | t | β | SE | t |
| Intercept | -0.65 | 0.73 | -0.89 | 5.50 | 0.09 | 60.40 | 5.50 | 0.10 | 56.27 | 5.49 | 0.15 | 36.47 | 5.65 | 0.20 | 28.34 |
| Type: Idiom | 0.61 | 0.22 | 2.79\*\* | -0.05 | 0.03 | -1.74+ | -0.06 | 0.03 | -1.74+ | -0.11 | 0.04 | -2.50\* | -0.18 | 0.06 | -3.07\*\* |
| Condition: Congruent | 0.18 | 0.25 | 0.72 | 0.01 | 0.04 | 0.35 | 0.01 | 0.04 | 0.28 | 0.03 | 0.06 | 0.57 | 0.01 | 0.07 | 0.14 |
| Condition: Swedish | -0.81 | 0.29 | -2.84\*\* | 0.04 | 0.04 | 1.22 | 0.05 | 0.04 | 1.22 | 0.07 | 0.05 | 1.27 | 0.14 | 0.07 | 2.07\* |
| Type\*Condition: Congruent | -0.39 | 0.31 | -1.29 | 0.01 | 0.04 | 0.18 | -0.01 | 0.03 | -0.19 | 0.00 | 0.06 | 0.01 | -0.01 | 0.08 | -0.07 |
| Type\*Condition: Swedish | -0.70 | 0.37 | -1.91\* | 0.06 | 0.04 | 1.58 | 0.11 | 0.04 | 2.36\* | 0.34 | 0.06 | 5.85\*\*\* | 0.52 | 0.08 | 6.74\*\*\* |
| Control predictors: |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| Word 2 Length | -0.18 | 0.07 | -2.56\* | -0.01 | 0.01 | -1.75+ | -0.00 | 0.01 | -0.43 | 0.00 | 0.01 | 0.34 | 0.00 | 0.02 | 0.02 |
| Word 2 Frequency (log) | 0.01 | 0.06 | 0.21 | -0.02 | 0.01 | -2.20\* | -0.02 | 0.01 | -2.23\* | -0.01 | 0.01 | -0.48 | 0.00 | 0.02 | 0.03 |
| Random effects | Variance | Variance | Variance | Variance | Variance |
| Item | 0.199 |  |  | 0.003 |  |  | 0.003 |  |  | 0.014 |  |  | 0.024 |  |  |
| Subject | 0.357 |  |  | 0.020 |  |  | 0.021 |  |  | 0.058 |  |  | 0.110 |  |  |
| Subject | Type | n/a |  |  | 0.001 |  |  | 0.001 |  |  | 0.001 |  |  | 0.006 |  |  |
| Subject | Condition: Congruent | n/a |  |  | 0.013 |  |  | 0.010 |  |  | 0.018 |  |  | 0.016 |  |  |
| Subject | Condition: Swedish | n/a |  |  | 0.008 |  |  | 0.012 |  |  | 0.007 |  |  | 0.013 |  |  |
| Residual | n/a |  |  | 0.078 |  |  | 0.094 |  |  | 0.161 |  |  | 0.275 |  |  |

*Table C.* Separate mixed effects model analysis for English native speakers, word level measures. For condition, native language is taken to be the baseline. Table displays coefficients (β), standard error (SE) and t-values (z-values for likelihood of skipping), with significance values estimated by the lmerTest package in R (version 2.0-11; Kuznetsova, Brockhoff & Christensen, 2014): \*\*\* p < .001, \*\* p < .01, \* p< .05