Additional Supplementary Material: Partial Analyses with All Participants (Outlier Candidates Included)

1. Semantically Meaningful Sentence Results

As shown in Table 1, the selected random effects structure was the maximal model (Model 1). The LRTs for fixed effects (Table 2) indicated significant main effects of Pro and L1. Post hoc contrasts for Pro showed significant differences for vowel vs. consonant and for vowel vs. vowel epenthesis on the odds scale (see Table 3). Evaluation of the no-interaction model (Model 4) indicated the following issues in the full model; however, visual inspection suggested they were minor: 1) the KS test was significant (p = .004), but the QQ plot aligned closely with the 45° line; 2) some deviations in within-group uniformity were observed, but the boxplots showed most values within the 0.25–0.75 band; and 3) Levene's test for homogeneity of variance was significant, although the boxplots were closely aligned and the differences appeared minor. The conditional R^2 and marginal R^2 were 0.661 and 0.047, respectively. Table 4 summarises the model output.

Table 1

Random Effects Model Comparison for Semantically Meaningful Sentences

Sample size	Total observations = 11200 Subjects = 100; Items = 112									
Selection process	**Model No.**	**Random effects**		**Model fit**			**Singular ?**	**LRT test**			
		Subject	**Items**	**AIC**	**BIC**	**logLik**		**Reference model**	**df**	**χ^2**	***p***
1. Building a maximal model											
	1	1 + Pro | sub	1 + L1 | item	8854	8963.8	-4412	No	-	-	-	-
2. Simplifying the model with LRT											
	2	1 + Pro | sub	1 | item	8901.5	8974.8	-4440.8	No	1	5	57.55	< .001
	3	1 | sub	1 + L1 | item	8861.1	8956.3	-4417.6	No	1	2	11.15	.004

Table 2

Fixed Effects Model Comparisons for Semantically Meaningful Sentences

Sample size		Total observations = 11200 Subjects = 100; Items = 112									
Model name	Model No.	Fixed effects	Random effects		Model fit			LRT test			
			Subjects	Items	AIC	BIC	logLik	Reference model	df	χ^2	*p*
1. Interaction model											
	1	Pro*L1	1 + Pro | sub	1 + L1 | item	8854	8963.8	-4412	-	-	-	-
2. No interaction model											
	4	Pro+L1	1 + Pro | sub	1 + L1 | item	8851.4	8946.6	-4412.7	1	2	1.40	.497
3. No Pro model											
	5	L1	1 + Pro | sub	1 + L1 | item	8854.9	8935.4	-4416.4	4	2	7.50	.024
4. No L1 model											
	6	Pro	1 + Pro | sub	1 + L1 | item	8859.1	8947	-4417.6	4	1	9.73	.002

Table 3

Pairwise Comparison Results

Contrast	Odds ratio	95 % CIs	Adjusted *p*
Vowel - Consonant	3.23	1.00 – 10.39	.049
Vowel - Vowel epenthesis	4.13	1.08 – 15.86	.036
Vowel epenthesis - Consonant	1.28	0.37 – 4.49	.888

Table 4

Results of Model 4 for Semantically Meaningful Sentences

	Fixed effects			Random effects	
				By subject	By items
Parameters	Estimate	*SE*	*z*	*SD*	*SD*
Intercept	2.45	0.41	6.04	0.82	2.30
Vowel - Consonant	-1.17	0.50	-2.35	0.49	-
Vowel - Vowel epenthesis	-1.42	0.57	-2.48	0.74	-
NES - NNES	-0.62	0.19	-3.20	-	0.41

2. Semantically Nonsensical Sentence Results

The maximal random effects model was over-parameterised in the subject-level random structure; accordingly, based on LRTs, Model 2 was selected (see Table 5). LRTs for the fixed effects (Table 6) indicated that none of the fixed effects reached statistical significance. Assessing the no-interaction model (Model 4) highlighted a few diagnostics in the full model, but visual checks suggested they were minor: 1) The KS test reached significance (p = .001); nevertheless, the QQ plot showed near-linear conformity with the 45° reference. Within-group uniformity exhibited slight irregularities, yet boxplots suggested only modest shifts within the 0.25–0.75 band. The conditional and marginal R^2 were 0.611 and 0.004, respectively. The model output is summarised in Table 7.

Table 5

Random Effects Model Comparison for Nonsensical Sentences

Sample size		Total observations = 11200 Subjects = 100; Items = 112									
Selection process	**Model No.**	**Random effects**		**Model fit**			**Singular ?**	**LRT Test**			
		Subjects	**Items**	**AIC**	**BIC**	**logLik**		**Reference model**	**df**	**χ^2**	***p***
1. Building a maximal model											
	1	1 + Pro | sub	1 + L1 | item	10223.4	10333.2	-5096.7	Yes	-	-	-	-
2. Simplifying the model with LRT											
	2	1 | sub	1 + L1 | item	10254.7	10327.9	-5117.4	No	-	-	-	-
	3	1 + Pro | sub	1 | item	10255	10328	-5117.4	No	2	2	9.98	.007

Table 6

Fixed Effects Model Comparison for Nonsensical Sentences

Sample size		Total observations = 11200 Subjects = 100; Items = 112									
Model name	Model No.	Fixed effects	Random effects		Model fit			LRT test			
			Subjects	Items	AIC	BIC	logLik	Reference model	df	χ^2	*p*
1. Interaction model											
	3	Pro*L1	1 + Pro | sub	1 | item	10255	10328	-5117.4	-	-	-	-
2. No interaction model											
	4	Pro+L1	1 + Pro | sub	1 | item	10255	10314	-5119.6	3	2	4.42	.110
3. No Pro model											
	5	L1	1 + Pro | sub	1 | item	10251	10295	-5119.7	4	2	0.29	.867
4. No L1 model											
	6	Pro	1 + Pro | sub	1 | item	10256	10307	-5121	4	1	2.94	.086

Table 7

Results of Model 4 for Nonsensical Sentences

	Fixed effects			Random effects	
				By subject	By items
Parameters	Estimate	*SE*	*z*	*SD*	*SD*
Intercept	-0.04	0.37	-0.10	0.64	2.23
Vowel - Consonant	-0.06	0.45	-0.13	-	-
Vowel - Vowel epenthesis	-0.25	0.50	-0.51	-	-
NES - NNES	-0.30	0.18	-1.73	-	0.35