Longitudinal effects of Context and Idiomaticity on the N400

Whereas we found a significant interaction between Context, Idiomaticity, and Correctness in the 300-400 ms time window (β = 1.05, *SE* = 0.34, *t* = 3.12, *p* = .002), the interaction just missed significance in the 400-500 ms time window (β = 0.70, *SE* = 0.36, *t* = 1.95, *p* = .051). However, as the interaction significantly improved model fit compared with models only including a two-way interaction in the 300-400 ms and 400-500 ms time window, we performed planned comparisons on the interaction between Context, Idiomaticity, and Correctness in both time windows.

With respect to the difference between the baseline and follow-up session we found a significant main effect for Session in the 300-400 ms (β = 0.19, *SE* = 0.09, *t* = 2.01, *p* = .044), as well as the 400-500 ms time window (β = 0.29, *SE* = 0.10, *t* = 2.91, *p* = .004), indicating that mean voltages were significantly more positive in the follow-up compared with the baseline session. In contrast to our predictions, Session did not interact with Context, Idiomaticity, or Correctness, meaning that the difference in mean voltages between the two sessions cannot be attributed to a longitudinal change in participants' brain responses to the experimental manipulation. A complete overview of the final models' coefficients can be found in **Table S1** in below.

Table S1. Coefficients of final models fitted to mean voltages in the 300-400 ms and 400-500 ms time window. Significant effects are printed in bold.

Fixed effects		*300-400 ms*				*400-500 ms*			
Factor		**Estimate**	**SE**	***t*-value**	**Pr(>|t|)**	**Estimate**	**SE**	***t*-value**	**Pr(>|t|)**
1	Intercept	**1.82**	**0.38**	**4.89**	**< .001**	**1.33**	**0.32**	**4.21**	**< .001**
2	Session (T2)	**0.19**	**0.09**	**2.01**	**.044**	**0.29**	**0.10**	**2.91**	**.004**
3	Context (predictive)	**0.55**	**0.20**	**2.77**	**.006**	**0.58**	**0.21**	**2.78**	**.007**
4	Idiomaticity (idiom)	**0.75**	**0.24**	**3.12**	**.003**	**0.73**	**0.26**	**2.80**	**.007**
5	Correctness (incorrect)	**-0.72**	**0.22**	**-3.22**	**.001**	**-1.27**	**0.24**	**-5.19**	**< .001**
6	Hemisphere (right)	-0.04	0.08	-0.47	.641	**-0.39**	**0.09**	**-4.38**	**< .001**
7	Context (predictive) * Idiomaticity (idiom)	**-0.71**	**0.24**	**-3.01**	**.003**	**-0.64**	**0.25**	**-2.51**	**.012**
8	Context (predictive) * Correctness (incorrect)	**-0.78**	**0.28**	**-2.80**	**.005**	**-0.66**	**0.30**	**-2.22**	**.027**
9	Idiomaticity (idiom) * Correctness (incorrect)	-0.33	0.24	-1.37	.171	**0.84**	**0.26**	**3.28**	**.001**
10	Context (predictive) * Idiomaticity (idiom) * Correctness (incorrect)	**1.05**	**0.34**	**3.12**	**.002**	0.70	0.36	1.95	.051

Random effects								
Groups		**Factor**	**Variance**	**SD**	**Corr.**	**Variance**	**SD**	**Corr.**
18	Subject	Intercept	2.77	1.66		1.65	1.29	
		Idiomaticity (idiom)	0.72	0.85	-0.28	0.86	0.93	-0.10
19	Target	Intercept	1.05	1.02		1.29	1.14	
		Context (predictive)	1.08	1.04	-0.53	1.10	1.05	-0.52
20	Residual		23.94	4.89		28.89	5.28	

Figures S1-S4 present participants' grand average ERPs in response to correct and incorrect literal and idiomatic sentences preceded by a neutral or predictive context in the 300-400 ms and 400-500 ms time window (green areas). **Figures S1 and S2** present mean voltages in the baseline session, whereas **Figures S3 and S4** present mean voltages in the follow-up session. Statistical analyses indicated that the effects did not significantly differ between the two sessions.

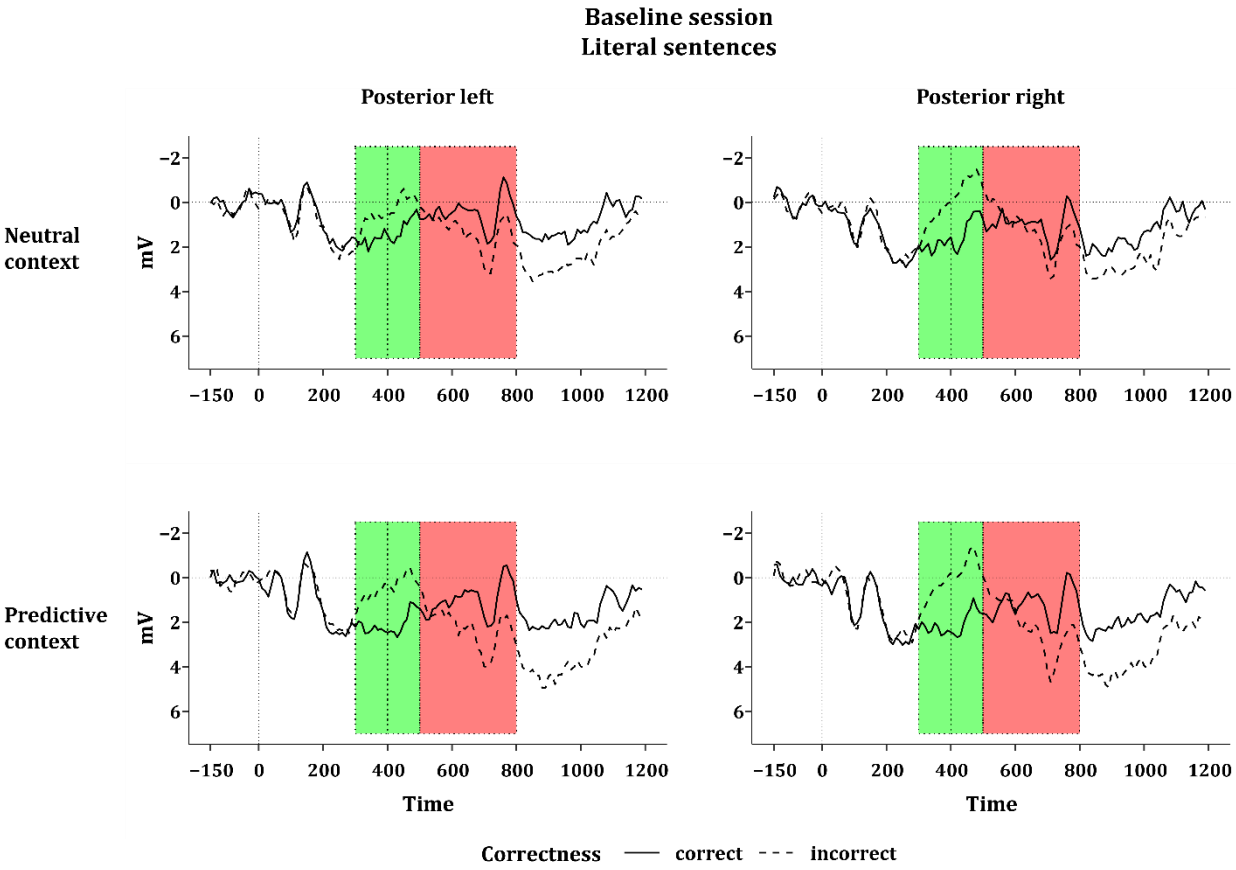

Figure S1. Mean voltages in the significant three-way interaction between Context, Idiomaticity, and Correctness in the 300-400 ms and 400-500 ms time window (green areas). Solid and dotted lines represent mean voltages elicited by correct and incorrect literal sentences in the baseline session. Negative values are plotted upwards.

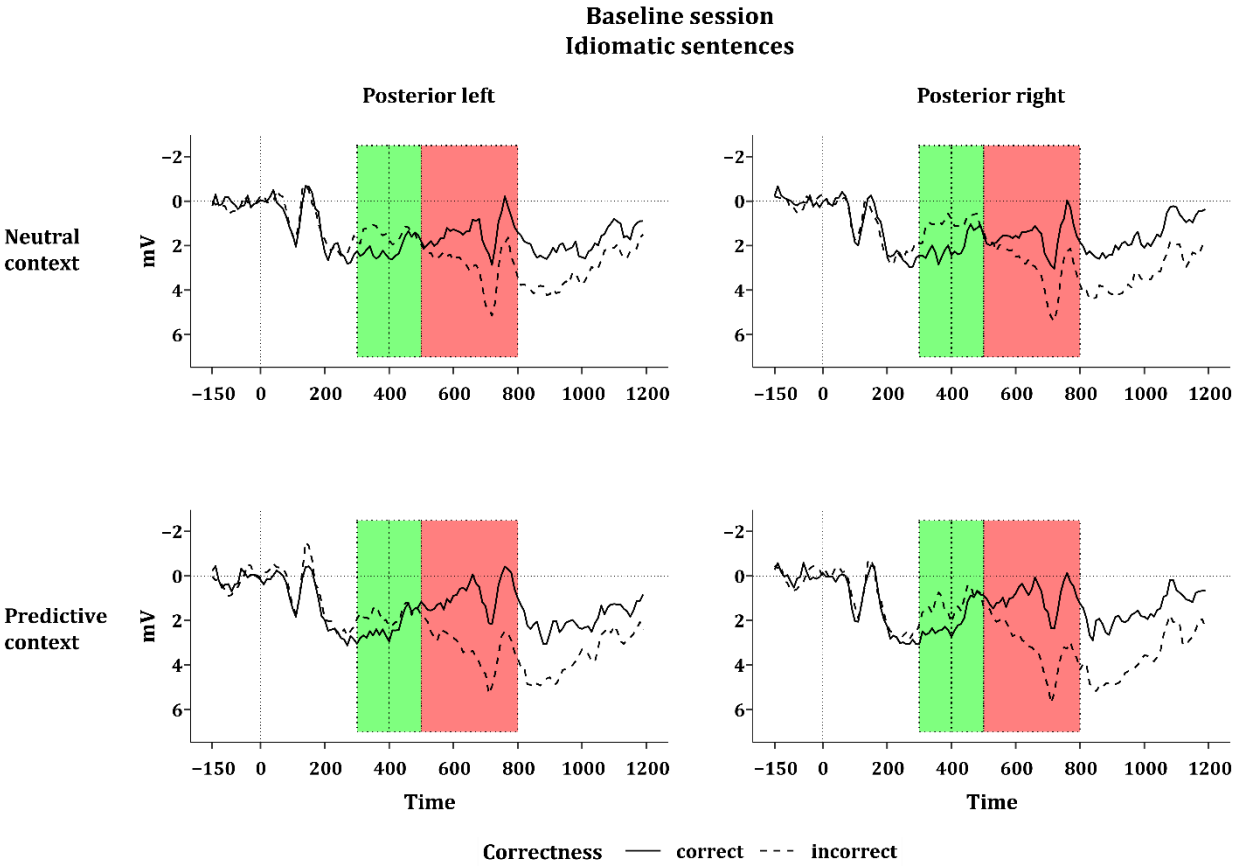

Figure S2. Mean voltages in the significant three-way interaction between Context, Idiomaticity, and Correctness in the 300-400 ms and 400-500 ms time window (green areas). Solid and dotted lines represent mean voltages elicited by correct and incorrect idiomatic sentences in the baseline session. Negative values are plotted upwards.

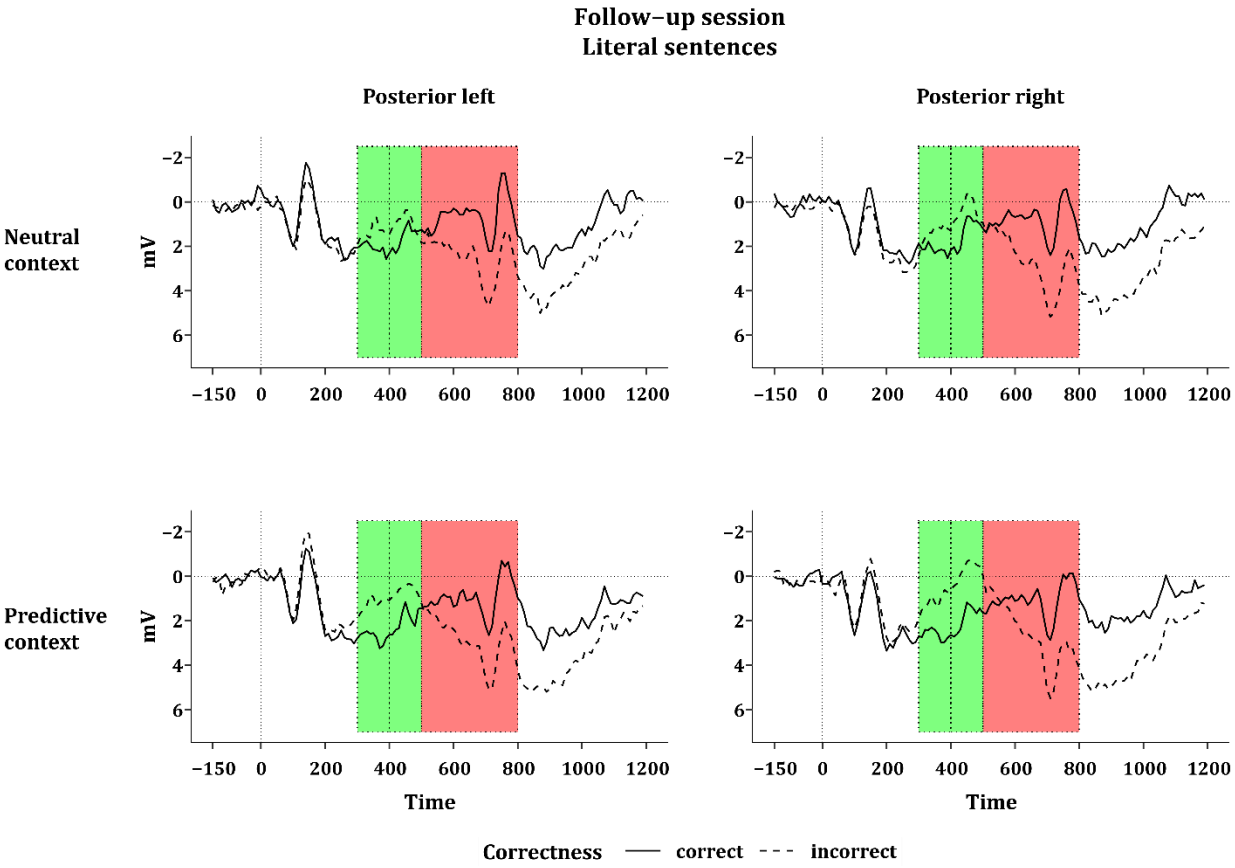

Figure S3. Mean voltages in the significant three-way interaction between Context, Idiomaticity, and Correctness in the 300-400 ms and 400-500 ms time window (green areas). Solid and dotted lines represent mean voltages elicited by correct and incorrect literal sentences in the follow-up session. Negative values are plotted upwards.

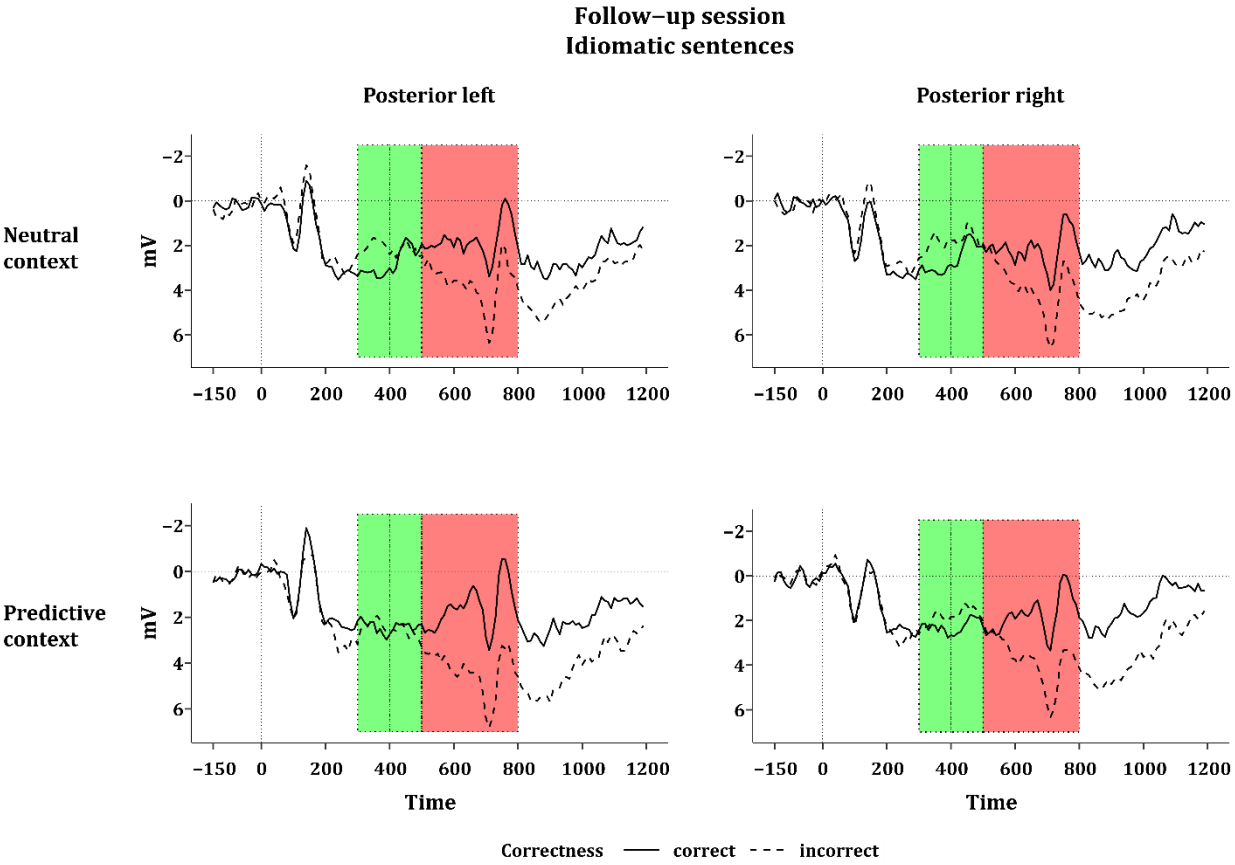

Figure S4. Mean voltages in the significant three-way interaction between Context, Idiomaticity, and Correctness in the 300-400 ms and 400-500 ms time window (green areas). Solid and dotted lines represent mean voltages elicited by correct and incorrect idiomatic sentences in the follow-up session. Negative values are plotted upwards.

Table S2 provides an overview of the planned comparisons that were carried out between mean voltages elicited by correct and incorrect target words in literal and idiomatic sentences preceded by a neutral or predictive context sentence. The results of the planned comparisons are summarized in **Figure S5**. As Session was not part of the interaction, mean voltages were averaged over the baseline and follow-up session, as well as over the left and right hemisphere.

With regard to the difference in mean voltages elicited by correct and incorrect target words in literal sentences we found significantly more negative mean voltages in response to incorrect relative to correct target words in literal sentences preceded by a neutral or predictive context sentence in the 300-400 ms as well as the 400-500 ms time window. This indicates that there was a significant N400 effect in literal sentences in both the baseline and follow-up. In idiomatic sentences we also found significantly more negative mean voltages in response to incorrect relative to correct

target words in both the neutral and predictive context condition. However, for idiomatic sentences the effect was only significant in the 300-400 ms time window.

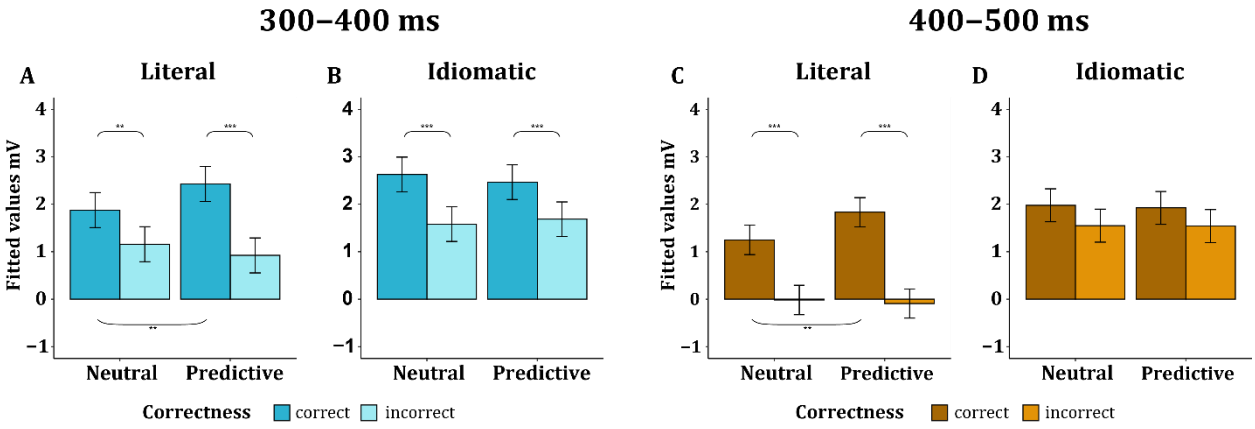

Figure S5. Mean voltages per condition illustrating the interaction Context * Idiomaticity * Correctness as fitted to the data in the 300-400 ms and 400-500 ms time window. Mean voltages are averaged over the baseline and follow-up session, because Session did not significantly interact with Context, Idiomaticity, and Correctness in the 300-400 ms and 400-500 ms time windows.

Regarding the effect of context on literal sentences we found that in both the 300-400 ms and 400-500 ms time window elderly adults' brain response was significantly less negative in response to literal sentences preceded by a predictive compared with a neutral context sentence. However, the effect was only significant for correct target words. In contrast, regarding the effect of context on idiomatic sentences we found that neutral versus predictive context sentences did not significantly affect mean voltages in response to correct or incorrect idiomatic sentences in the 300-400 ms, nor the 400-500 ms time window. This finding is inconsistent with La Roi et al. (2020), who found a context effect on the N400 elicited by both literal and idiomatic sentences. The discrepancy between the results in the present study and that by La Roi et al. (2020) could arise from the fact that La Roi et al. measured N400 amplitudes in both younger adults and elderly adults. The N400 is found to be generally larger in younger compared with elderly adults (e.g., Kutas & Iragui, 1998; Wlotko, Lee, & Federmeier, 2010) and younger adults generally show stronger context effects in the N400 (e.g., Dave et al., 2018; Federmeier & Kutas, 2005; Payne & Silcox, 2019). The strong context effects on younger

adults' N400 for both literal and idiomatic sentences could have masked the fact that elderly adults only use context for the processing of literal sentences. As the present study only measured the N400 in elderly adults, subtle context effects are more likely to become visible. Thus, one of the strengths of investigating brain responses longitudinally in one and the same group is that this approach may be more suitable to detect subtle intraindividual differences in language processing over time.

With respect to the difference between literal and idiomatic sentences we found that in both the 300-400 ms and 400-500 ms time window, correct literal sentences preceded by a neutral context elicited significantly more negative mean voltages than correct idiomatic sentences. Furthermore, incorrect literal sentences preceded by a neutral context elicited significantly more negative mean voltages than incorrect idiomatic sentences, but only in the 400-500 ms time window. For sentences preceded by a predictive context we found a slightly reversed pattern: incorrect literal sentences elicited significantly more negative mean voltages than incorrect idiomatic sentences, but not when the target word was correct. However, the effect was only significant in the 400-500 ms time window.

Longitudinal effects of Context and Idiomaticity on the P600

The four-way interaction that was included in the final model fitted on the mean voltages in the 500-800 ms time window did not reach significance itself (β = 0.68, *SE* = 0.72, *t* = 0.95, *p* = .343). However, a significant underlying interaction between Session and Correctness (β = 1.04, *SE* = 0.36, *t* = 2.90, *p* = .004) showed that participants' brain responses to violations differ between the baseline and follow-up session, consistent with the findings from the analysis on general longitudinal changes in the P600 effect. Furthermore, Idiomaticity significantly interacted with Correctness (β = 0.72, *SE* = 0.32, *t* = 2.25, *p* = .025), illustrating that brain responses to violations depend on whether the target is presented in a literal or idiomatic test sentence. In addition, the significant interaction between Context and Idiomaticity (β = -1.06, *SE* = 0.32, *t* = -3.32, *p* < .001) indicated that the difference between mean voltages elicited by literal versus idiomatic sentences depends on the type of

preceding context sentence. A complete overview of the final model's coefficients can be found in **Table S3** in the supplementary materials.

As the interaction between Session and Correctness has already been discussed in the paper, here we only elaborate on the interaction between Idiomaticity and Correctness and between Context and Idiomaticity. **Figures S1-S4** summarize the effects of Context, Idiomaticity, and Correctness in four plots presenting participants' grand average ERPs in response to correct and incorrect literal and idiomatic sentences preceded by a neutral or predictive context in the 500-800 ms time window (red area). **Figures S1 and S2** present mean voltages in the baseline session, whereas **Figures S3 and S4** present mean voltages in the follow-up session. Statistical analyses indicated that the effects did not significantly differ between the two sessions.

Table S2. Planned comparisons on the levels of the significant interaction between Context, Idiomaticity, and Correctness in the 300-400 ms time window. Significant effects are printed in bold.

		Contrast	300-400 ms				400-500 ms			
			β	SE	*z*-value	*p*	β	SE	*z*-value	*p*
Neutral context	Literal	Incorrect - Correct	**-0.72**	**0.22**	**-3.22**	**.001**	**-1.27**	**0.24**	**-5.19**	**< .001**
	Idiomatic	Incorrect - Correct	**-1.04**	**0.22**	**-4.67**	**< .001**	-0.43	0.24	-1.75	.080
	Correct	Literal - Idiomatic	**-0.75**	**0.24**	**-3.12**	**.001**	**-0.73**	**0.26**	**-2.80**	**.005**
	Incorrect	Literal - Idiomatic	-0.42	0.24	-1.76	.078	**-1.57**	**0.26**	**-5.99**	**< .001**
Predictive context	Literal	Incorrect - Correct	**-1.50**	**0.22**	**-6.79**	**< .001**	**-1.92**	**0.24**	**-8.10**	**< .001**
	Idiomatic	Incorrect - Correct	**-0.78**	**0.22**	**-3.53**	**< .001**	-0.38	0.24	-1.61	.108
	Correct	Literal - Idiomatic	-0.04	0.24	-0.16	.873	-0.09	0.26	-0.35	.729
	Incorrect	Literal - Idiomatic	**-0.76**	**0.24**	**-3.16**	**.002**	**-1.63**	**0.26**	**-6.25**	**< .001**
Literal sentences	Correct	Neutral - Predictive	**-0.55**	**0.20**	**-2.77**	**.006**	**-0.58**	**0.21**	**-2.78**	**.006**
	Incorrect	Neutral - Predictive	0.24	0.20	1.18	.237	0.08	0.21	0.36	.719
Idiomatic sentences	Correct	Neutral - Predictive	0.16	0.20	0.81	.416	0.05	0.21	0.26	.795
	Incorrect	Neutral - Predictive	-0.10	0.20	-0.52	.607	0.01	0.21	0.04	.969

Table S3. Coefficients of final model fitted to mean voltages in the 500-800 ms time window. Significant effects are printed in bold.

Fixed effects		*500-800 ms*			
Factor		**Estimate**	**SE**	***t*-value**	**Pr(>|t|)**
1	Intercept	0.57	0.33	1.73	.089
2	Session (T2)	0.24	0.26	0.93	.354
3	Context (predictive)	**0.54**	**0.26**	**2.07**	**.039**
4	Idiomaticity (idiom)	**0.97**	**0.28**	**3.45**	**< .001**
5	Correctness (incorrect)	**0.71**	**0.28**	**2.49**	**.013**
6	Hemisphere (right)	0.10	0.09	1.13	.259
7	Category fluency	**0.05**	**0.02**	**2.98**	**.003**
8	Session (T2) * Context (predictive)	-0.45	0.36	-1.27	.204
9	Session (T2) * Idiomaticity (idiom)	0.28	0.36	0.78	.435
10	Context (predictive) * Idiomaticity (idiom)	**-1.06**	**0.32**	**-3.32**	**< .001**
11	Session (T2) * Correctness (incorrect)	**1.04**	**0.36**	**2.90**	**.004**
12	Context (predictive) * Correctness (incorrect)	0.30	0.37	0.80	.422
13	Idiomaticity (idiom) * Correctness (incorrect)	**0.72**	**0.32**	**2.25**	**.025**
14	Session (T2) * Context (predictive) * Idiomaticity (idiom)	0.42	0.51	0.84	.403
15	Session (T2) * Context (predictive) * Correctness (incorrect)	-0.51	0.51	-1.01	.311
16	Session (T2) * Idiomaticity (idiom) * Correctness (incorrect)	-0.82	0.51	-1.62	.105
17	Context (predictive) * Idiomaticity (idiom) * Correctness (incorrect)	0.33	0.46	0.72	.470
18	Session (T2) * Context (predictive) * Idiomaticity (idiom) * Correctness (incorrect)	0.68	0.72	0.95	.343

Random effects					
Groups		**Factor**	**Variance**	**SD**	**Corr.**
18	Subject	Intercept	1.64	1.28	
		Idiomaticity (idiom)	0.67	0.82	-0.08
19	Target	Intercept	1.37	1.17	
		Context (predictive)	1.68	1.29	-0.59
20	Residual		26.65	5.16	

Table S4 gives an overview of the planned comparisons that were carried out on the interactions between Session and Correctness, between Idiomaticity and Correctness, and between Context and Idiomaticity. The results of the planned comparisons are summarized in **Figure S6**. Below, we discuss the findings on the interaction between Idiomaticity and Correctness and between Context and Idiomaticity. To carry out the planned comparisons, mean voltages were averaged over the left and right hemisphere and the baseline session and the follow-up session, as well as over neutral and predictive contexts (for the interaction between Idiomaticity and Correctness) or over correct and incorrect sentences (for the interaction between Context and Idiomaticity).

Table S4. Planned comparisons between the levels of the significant interactions between Context and Idiomaticity, between Session and Correctness, and between Idiomaticity and Correctness in the 500-800 ms time window. Significant effects are printed in bold.

Contrast		500-800 ms			
		β	SE	*z*-value	*p*
Baseline	Incorrect - Correct	**1.30**	**0.18**	**7.31**	**< .001**
Follow-up	Incorrect - Correct	**1.84**	**0.20**	**9.44**	**< .001**
Correct	Baseline - Follow-up	-0.26	0.14	-1.94	.052
Incorrect	Baseline - Follow-up	**-0.80**	**0.14**	**-5.91**	**< .001**
Literal	Incorrect - Correct	**1.25**	**0.19**	**6.67**	**< .001**
Idiomatic	Incorrect - Correct	**1.90**	**0.19**	**10.13**	**< .001**
Correct	Literal - Idiomatic	**-0.68**	**0.21**	**-3.25**	**.001**
Incorrect	Literal - Idiomatic	**-1.33**	**0.21**	**-6.30**	**< .001**
Neutral	Literal - Idiomatic	**-1.27**	**0.21**	**-6.01**	**< .001**
Predictive	Literal - Idiomatic	**-0.75**	**0.21**	**-3.54**	**< .001**
Literal	Neutral - Predictive	**-0.34**	**0.16**	**-2.15**	**.032**
Idiomatic	Neutral - Predictive	0.18	0.16	1.13	.258

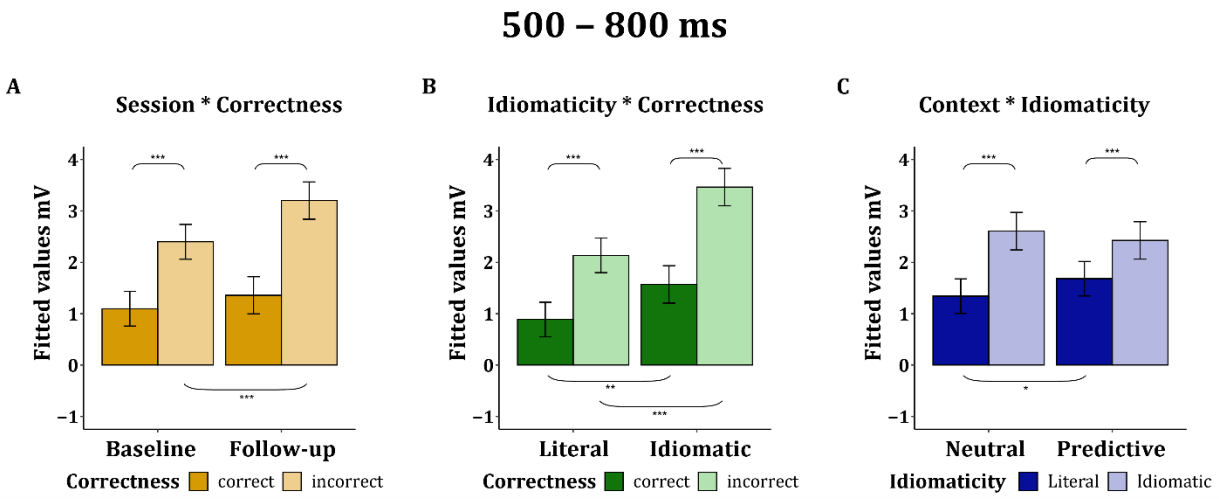

Figure S6. Fitted mean voltages per condition illustrating the interaction Context * Idiomaticity * Correctness as fitted to the data in the 500-800 ms time window. Mean voltages are averaged over the predictors that are not included in the interaction.

Direct comparison of mean voltages elicited by correct and incorrect target words showed that incorrect target words evoked significantly more positive voltages than correct target words in both literal and idiomatic sentences. With respect to the difference between literal and idiomatic sentences planned comparisons showed that mean voltages were significantly more positive in response to idiomatic than literal sentences in both the correct and the incorrect target word condition.

Regarding the effect of context we found that mean voltages were significantly more positive for literal sentences preceded by predictive compared with neutral contexts. However, mean voltages in response to idiomatic sentences were not significantly affected by predictive or neutral contexts. Furthermore, idiomatic sentences evoked significantly more positive mean voltages than literal sentences in both types of context.