APPENDIX S1

The following text provides a description of (i) the statistical models that were employed (viz., generalised linear mixed-effects regression on untransformed response times), (ii) the assessment of the models' random structure, and (iii) the software versions used for data analysis, as well as the specific arguments to the function call.

1. Analysis of untransformed RTs with generalised linear models

As recommended by Lo and Andrews (2015), in the current paper we analysed the response time (RT) data with generalised linear mixed-effects regression, by including in the models the assumption that the data followed an 'RT-like' distribution—specifically, an inverse Gaussian.

As RT distributions are heavily skewed, it is common in experimental psycholinguistics to analyse the logarithm or the reciprocal of RTs, so that assumptions regarding the normality of residuals can be satisfied. This was also the approach taken in Farhy, Veríssimo, and Clahsen (in press), in which we analysed the reciprocal of RTs (-1000/RT), rather than the actual times produced in the experiment. However, concerns have been raised about the analysis of transformed RTs being associated with serious problems (Balota, Aschenbrenner, & Yap, 2013; Lo & Andrews, 2015; O'Malley & Besner, 2013). In particular, Balota et al. have demonstrated that nonlinear transformations (such as the reciprocal) may give rise to spurious interactions by distorting purely additive relationships. A solution, recently proposed by Lo and Andrews, is to employ generalised linear mixed-effects regression. In this approach, raw (untransformed) RTs are directly analysed, but at the same time, it is possible to include the assumption that the data follows a skewed distribution (see Lo & Andrews, for further details; for recent examples of this type of analysis, see, e.g., Masson, Rabe, & Kliegl, 2017; Medeiros & Duñabeitia, 2016). An important additional benefit of this approach is that effects can be readily interpreted in their true scale, that is, every estimate is expressed as a difference in milliseconds.

2. Random structure of the statistical model

In the current study, the data were analysed with regression models with crossed random effects for participants and items (Baayen, Davidson, & Bates, 2008). As recommended by Matuschek, Kliegl, Vasishth, Baayen, and Bates (2017), random slopes for the different predictors were tested for inclusion on the basis of the models' AIC, a measure of goodness of fit. Against a simple, intercept-only between-group regression model (with categorical fixed effects Prime Type, Form Type, and Group), we tested all possible random slopes individually and obtained the AIC of the resulting models. A random by-item slope for Group (L1, L2) improved fit the most (i.e., led to the lowest AIC), for both the RT model and the accuracy model. Further inclusion of additional random slopes led to models that did not converge or did not improve fit. Additional follow-up analyses (within each group) were conducted with intercept-only models, as Group was not a predictor in these models.

3. Software versions and function call

Data were analysed using the *lme4* package (version 1.1-12; Bates, Maechler, Bolker, & Walker, 2015) for the R language (version 3.4.1). Specifically, in the case of the RT analysis, models were fit using the following function call:

```
glmer(…, family=inverse.gaussian(link="identity"),
glmerControl(optimizer="bobyqa", optCtrl=list(maxfun=30000)))
```

In the case of the accuracy analysis, the following function call was used:

```
glmer(…, family=binomial, glmerControl(optimizer="bobyqa",
optCtrl=list(maxfun=30000)))
```

References

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, *59*, 390–412.

Balota, D. A., Aschenbrenner, A. J., & Yap, M. J. (2013). Additive effects of word frequency and stimulus quality: The influence of trial history and data transformations. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *39*, 1563–1571.

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*, 1–48.

Farhy, Y., Veríssimo, J., & Clahsen, H. (in press). Universal and particular in morphological processing: Evidence from Hebrew. *The Quarterly Journal of Experimental Psychology*. Advance online publication. doi:10.1080/17470218.2017.1310917

Lo, S., & Andrews, S. (2015). To transform or not to transform: Using generalized linear mixed models to analyse reaction time data. *Frontiers in Psychology*, *6 :1171*.

Masson, M. E., Rabe, M. M., & Kliegl, R. (2017). Modulation of additive and interactive effects by trial history revisited. *Memory & Cognition*, *45*, 480–492.

Matuschek, H., Kliegl, R., Vasishth, S., Baayen, H., & Bates, D. (2017). Balancing Type I error and power in linear mixed models. *Journal of Memory and Language*, *94*, 305–315.

Medeiros, J., & Duñabeitia, J. A. (2016). Not everybody sees the ness in the darkness: Individual differences in masked suffix priming. *Frontiers in Psychology*, *7:1585*.

O'Malley, S., & Besner, D. (2013). Reading aloud: Does previous trial history modulate the joint effects of stimulus quality and word frequency? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *39*, 1321–1325.