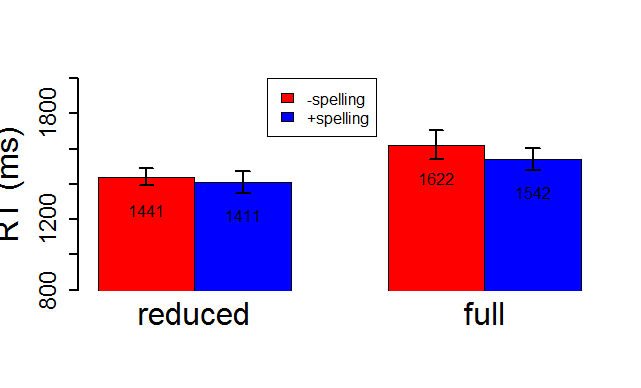


Figure



Figure

### Experiment 1

***Results***

*Vocabulary test*

The overall hit rate in the vocabulary test was 86.0%. Participants in the +spelling group identified the correct Dutch translations in 84.2% of the cases, and in the –spelling group in 87.7% of the cases. According to a chi-squared test for independence, there was no correlation of hit rates and *Spelling* group(χ2 = 2.82; df = 1; ptwo-tailed= .09).

*Lexical decision*

Three word types that elicited an error rate higher than 40 % in the reduced condition were excluded.

Subsequently, data from one participant with an overall error rate higher than 40 % were excluded. In sum, the analyses of response accuracies are based on 1113 trials.

For the analyses of RTs, outliers were removed after reaction times had been log-transformed. Outliers were defined as those responses with reaction times that were more than 2.5 standard deviations away from the mean reaction time in the target conditions. This resulted in the rejection of 32 trials. Subsequently, all 305 *nonword* responses were removed. In sum, the analyses of RTs are based on 776 trials.

##### Response accuracy

Response accuracy as a function of *Reduction Type* and *Matching* is plotted in Figure 1. The influence of the predictors *Reduction* and *Spelling* and their interaction on response accuracy was analyzed

Table 1. *Statistical analysis of response accuracy in Experiment 1. Reference levels – Reduction: reduced, Spelling:   
-Spelling*

Fixed effects:

β Std.Error Z p

(Intercept) 2.3259 0.3005 7.741 9.90e-15 \*\*\*

Reduction -2.3891 0.3164 -7.551 4.33e-14 \*\*\*

spelling -0.8656 0.3487 -2.482 0.013 \*

Reduction:Spelling 2.0977 0.3203 6.549 5.77e-11 \*\*\*

using linear mixed effect models with a binomial link function. We included random intercepts for *Participant* and *Word Type* and random slopes for *Reduction* by *Word Type* in the final model. The final model is summarized in Table 1.

The model indicates that exposure to spelling significantly decreased the number of correct responses to reduced variants. Moreover, the effect of *Spelling* significantly varied as a function of *Reduction*, increasing the number of word responses to full variants. In order to determine whether this effect on responses to full variants was significant, we reran the model with full variants on the intercept. With this model, the effect of *Spelling* turned out significant, as well (β = 1.23, p < .0001).

##### RTs

Mean RT measured from word onset as a function of *Matching* and *Reduction Type* is plotted in Figure 2. Prior to the statistical analyses, RTs were log-transformed in order to approximate a normal distribution of RTs.

We included three additional control variables in the models: log-*Duration*, *log-RT* *of the* *Preceding Trial, Trial Number*. The final statistical model was obtained using a backwards stepwise selection procedure, in which insignificant predictors were removed from the model. Response times that were more than 2.5 standard deviations away from the residuals predicted by the model were removed. Subsequently, the model was refit. As for the statistical analyses of response accuracy, random intercepts for *Word Type* and *Participant* were included in the model. Moreover random slopes for *Reduction* by *Word Type* and *Participant* were included. The final model is summarized in Table 2.

Table 2. *Statistical analysis of RTs in Experiment 1. Reference levels – Reduction: reduced*

Fixed effects:

Estimate Std. Error t value

(Intercept) 5.9300038 0.4467900 13.272

Reduction 0.0951284 0.0254686 3.735

Trial Number -0.0007873 0.0001257 -6.265

Log-Duration 0.1382464 0.0609352 2.269

Preceding Log-RT 0.0604272 0.0211671 2.855

A significant effect of *Reduction* indicates that RTs were significantly longer for reduced variants. The significant effects of *Trial Number*, *Log-Duration*, and *Log-RT of the Preceding Trial* indicate that RTs were significantly longer, at later trials, for longer target variants, and varied as a function of the RTs of the preceding trials. Crucially, there was no effect of *Spelling* on RTs.

### Experiment 2

***Results***

*Vocabulary test*

The overall hit rate in the vocabulary test was 87.7%. There was no correlation of hit rates and *Spelling* group(+spelling: 87.8%, -spelling: 87.5%; χ2 = 0.01; df = 1; ptwo-tailed= .93).

*Lexical decision*

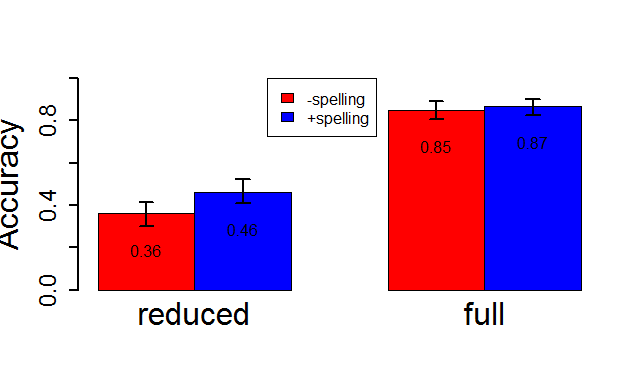
Four participants with an error rate above 40 % in all 192 trials were excluded (115 trials). The analyses of accuracy data are therefore based on 1200 trials.

For the analyses of RTs, 32 outlier trials that were more than 2.5 standard deviations away from the mean RT in the target conditions were removed. Moreover, all 416 *nonword* responses were removed. Consequently, analyses of RTs are based on 752 trials.

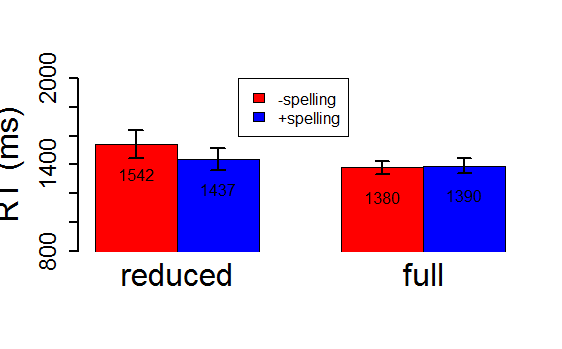
*Response accuracy*

Accuracy as a function of *Reduction* and Spelling is plotted in Figure 3.

As for Experiment 1, we analyzed the influence of *Reduction* and *Spelling* and their interaction on response accuracy. We obtained the final statistical model by using a backwards stepwise selection procedure, in which insignificant predictors were removed from the model. Random intercepts for *Word Type* and *Participant* were included in the final model, as well as random slopes for *Reduction*



Figure



Figure

by *Participant*. The only significant effect was that of *Reduction* (β = 3.14, p < .0001), indicating that full variants were recognized significantly more accurately than reduced variants.

*RTs*

Mean RT measured from word onset as a function of *Spelling* and *Reduction* is plotted inFigure 4**.**

We included log-*Duration* and *log-RT* *of the* *Preceding Trial* andTrial Number as additional control variables. Response times that were more than 2.5 standard deviations away from the residuals predicted by the model were removed. Subsequently, the model was refit. Random intercepts for *Word Type* and *Participant*, as well as random slopes for *Reduction* by *Word Type*, and random slopes for *Reduction* and *log-Duration* by *Participant* were included in the final model. The final statistical model is summarized in Table 3.

The model indicates that RTs were significantly longer for reduced variants, for longer words, and varied as a function of RTs of the preceding trials. Moreover, full variants were identified significantly faster than reduced variants when exposure to spelling had not been part of the training. There was no significant effect of spelling on reduced variants. However, the effect of *Spelling* significantly varied as a function of *Reduction*, leading to shorter RTs for reduced variants,

Table 3. *Statistical analysis of RTs in Experiment 2. Reference levels – Reduction: reduced, Spelling:   
-Spelling*

Fixed effects:

Estimate Std. Error t value

(Intercept) 4.78126 0.52790 9.057

Reduction -0.16821 0.03159 -5.324

Spelling -0.05362 0.04502 -1.191

Log-Duration 0.29667 0.07297 4.066

Preceding log-RT 0.07805 0.02540 3.073

Reduction:Spelling 0.07598 0.03771 2.015

and to longer RTs for full variants. In order to determine whether the effect of *Spelling* on RTs to full variants was significant, we reran the final model with full variants on the intercept. With that model, the effect of *Spelling* remained insignificant (β = 0.02, ns.).

### Experiment 3

***Results***

*Vocabulary test*

The overall hit rate in the vocabulary test was 85.2%. There was no correlation of hit rates and *Spelling* group(+spelling: 86.6%, -spelling: 83.8%; χ2 = 1.76; df = 1; ptwo-tailed= .18).

*Lexical decision*

Two words elicited an error rate above 40 % and were excluded (108 trials). Subsequently, data from two participants with an overall error rate above 40 % in all 192 trials were removed. The analyses of accuracy data are therefore based on 1144 trials.

For the analyses of RTs, 38 outlier trials were removed, and subsequently all 199 nonword *responses*. Analyses of RTs are therefore based on 907 trials.

*Response accuracy*

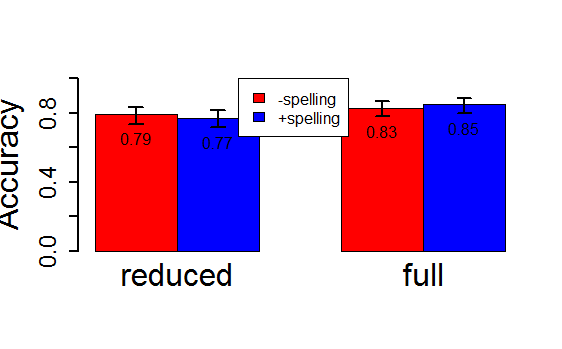
Response Accuracy as a function of *Spelling* and *Reduction* is plotted in Figure 5**.**

As for the preceding two experiments, we obtained the final statistical model using a backwards stepwise selection procedure, removing insignificant predictors at each step. Random intercepts for *Word Type* and *Participant* were included in the final model. The only significant predictor was *Reduction* (β = 0.49, p < .01) with full variants being recognized significantly more accurately.

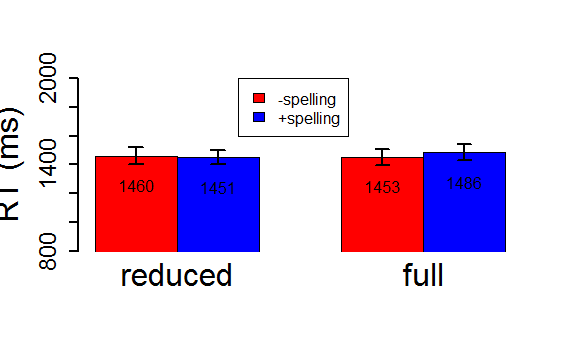
RTs

Mean RT measured from word onset as a function of *Spelling* and *Reduction* is plotted inFigure 6**.**

The final model is summarized in Table 4. RTs became faster at later trials, and they varied as function of word duration and RTs of the preceding trials. Crucially, no predictor of interest yielded a significant effect.



Figure



Figure

Table 4. *Statistical analysis of RTs in Experiment 3.*

Fixed effects:

Estimate Std. Error t value

(Intercept) 5.4486230 0.5688087 9.579

Trial Number -0.0003279 0.0001285 -2.553

log(duration) 0.1923124 0.0811736 2.369

previous\_log\_rt 0.0725720 0.0202846 3.578