Appendix S1

Control experiment: Native Dutch lexical decision in prime-target word pairs

Method

Participants

In total, 35 students from the Radboud University Nijmegen (31 females, mean age=23.17) participated. All were right-handed native speakers of Dutch and had normal or corrected to normal vision. Participation in the experiment was voluntary and compensated with course credits or a gift card.

Materials and Design

Materials were adapted from Experiment 1. Each idiom-final word was taken separately as a prime in a prime-target pair. For example, the word 'pols' ('wrist') was isolated from the previously exemplified idiom *Hij deed iets uit de losse pols* and was presented alone as a prime for each target word (e.g., 'pols – MAKKELIJK', 'wrist – EASY' for the FIG condition).

Procedure

The same procedure as in Experiment 1 was employed, except for the reduction of presentation of stimulus materials to single word prime-target pairs.

Results

Reaction times

Target words belonging to the expression *tussen twee vuren zitten*, which was excluded from Experiment 1, were also excluded here to maximize comparability between experiments. First, two participants were excluded for overall slow RTs at 2.5 SDs below the mean of all

participants. One participant was excluded for more than 20% errors. After participant exclusions, three items were excluded for overall outlier slow RTs (‘INGEWANDEN’ / ‘INTESTINES’, ‘KRUK’ / ‘STOOL', and ‘WIJD’ / ‘WIDE’) at 2.5 SDs below the overall mean. Outlier datapoints were then removed at both subject and item level above 2.5 SDs from the mean. One further participant and one further item (‘KAAK’ / ‘JAW’) were excluded for over 20% data loss after outlier analysis. Paired t-tests showed that targets remained balanced across conditions in terms of word length and word frequency after exclusions. Mean RTs are shown in Table 4 along with error rates.

Table S1.1. *Means and standard deviations for each Target Word Condition in the single prime-target experiment.*

	Figuratively related	**Literally related**	**Unrelated**
Mean RT (SD)	519 (101)	504 (97)	519 (104)
Error Rate (SD)	.04 (.05)	.05 (.06)	.03 (.05)

A linear mixed effects regression model was built by iteratively adding predictors and testing each model against its predecessor in an ANOVA until the most complex, theoretically relevant model had been reached. Idiom-level predictors such as transparency were not included as there was no idiom context. The final model took the log-transformed RTs as the dependent variable and included random slopes for Participant (over Trial Number), and Item. Target Word Frequency, Trial Number, and Target Word Length were included as independent predictors. A two-way interaction between Target Word Condition and Idiom-Final Word Frequency was included (for continuity, we will refer to the prime as the idiom-final word, even though there is no idiomatic context). Again, $t>1.96$ was taken as convention for interpreting statistical significance and releveling was applied to the model to compare

conditions. Results are summarized in Table 5 for the relevel of the model with the literal (LIT) condition on the intercept as this relevel provides the most telling contrasts for this single word prime-target experiment.

Table S1.2. *Releveled linear mixed effects regression model for Dutch lexical decision by Dutch L1 speakers with single word-target primes with literal (LIT) condition on the intercept.*

	Estimate	**SE**	**df**	**t-value**	**p-value**
LIT vs. FIG	.04325	.009615	1761	4.498	.000
LIT vs. UNREL	.03180	.009297	1858	3.420	.000
Target Word Frequency	-.04734	.009243	155.9	-5.122	.000
Target Word Length	-.008769	.003164	99.8	-2.771	.007
Idiom-Final Word Frequency	.01251	.01383	67.90	.904	.369
Trial Number	-.009431	.006154	28.50	-1.532	.136
(Condition 'LIT') FIG*Idiom-Final Word Frequency	-.03209	.01576	1858	-2.037	.042
(Condition 'LIT') UNREL*Idiom-Final Word Frequency	-.01153	.01566	1849	-.736	.462

Comparing target conditions showed that LIT targets were responded to faster than target words that were originally figuratively related to the idiom, e.g., FIG targets (Estimate=.04325, SE=.009615, t(1761)=4. 498, p<.001). They were also responded to faster than unrelated targets (Estimate=.03180, SE=.009297, t(1858)=3. 420, p<.001). However, there was no RT difference between FIG and LIT targets. A simple effect of Target Word Length reflected faster RTs to shorter targets across all conditions (Estimate=-.008769, SE=.003164, t(99.8)=-2.771, p<.01). A simple effect of Target Word Frequency indicated that

a higher target word frequency resulted in faster RTs across all experimental conditions (Estimate=-.04734, SE=.009243, t(155.9)=-5.122, p<.001). Furthermore, we found a Target Word Condition*Idiom-Final Word Frequency interaction effect when comparing slopes for this interaction between the FIG and the LIT conditions. The effect of Idiom-Final Word Frequency on RTs to FIG targets differed significantly from that LIT targets (Estimate=-.03209, SE=.01576, t(1858)=2.037, p<.05). This was the only significant difference in the interaction effects (see Figure 2).

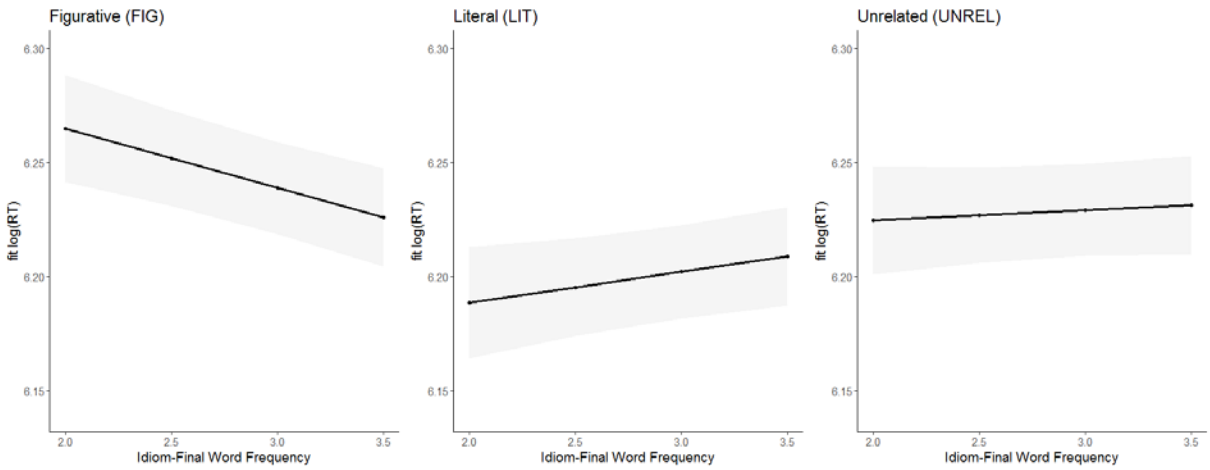

Figure S1.1. Idiom-Final Word Frequency effects for all three conditions (Figurative (FIG), Literal (LIT), and Unrelated (UNREL)) in the single word prime-target experiment.

Error analysis

Error rates in the single word prime-target lexical decision task averaged .05 across conditions with a maximum of .24 (SD=.05). Table 3 lists means and SDs for the accuracy analysis. The mean error rate for nonwords was .07 (SD=.05). A binary logistic regression using the same model as the RT data did not yield differences in error rates between experimental conditions.

Discussion

In this control experiment, we stripped idiom-final words of their idiomatic context and presented them in isolation in a prime – target setting (e.g., tand – kaak, translated: tooth - jaw). We obtained facilitatory priming effects in the LIT condition, indicating that our literally related target words were, indeed, semantically related to the idiom-final words. Furthermore, RTs on literally related target words were facilitated in comparison to both targets originally figuratively related to idioms, and unrelated targets. FIG and UNREL targets were not primed, which excludes the possibility for an arbitrary relationship between idiom-final nouns and figuratively related targets, therefore validating our stimulus materials and results found in Experiment 1.

Higher target word frequency was found to facilitate RTs similarly in all three target words conditions. A higher frequency of the figuratively, literally, or unrelated word resulted in faster RTs on the lexical decision.

We found that the effect of idiom-final word frequency differed between the FIG and LIT conditions especially. For the FIG condition, we found that a higher prime word frequency facilitated RTs in comparison to the LIT condition, as there was no effect of prime word frequency in the UNREL condition. We offer an interesting but admittedly speculative interpretation for these results in the form of two-step priming. This explanation assumes that the presented individual word and the idiom as a whole resonate, thus strengthening each other. This assumption is in line with hybrid models of idiom processing such as Sprenger et. al. (2006). In such models, activation spreads in two directions between lemma and conceptual level representations of the idiom via a meaning relationship between the two levels of representation. For instance, the item ‘tooth’ activates both semantically related words such as ‘jaw’, as well as idioms in which this item is contained. All idioms in this experiment were well-known to Dutch native speakers. In this way, participants may show a slight sensitivity to prime frequency in the figurative condition.