**Supplementary Materials: Focal expressions and data**

These materials consist of conventional multiword expressions (MWEs) and various associated measures. All these MWEs are composed of two words. The most common structures are N-N, Adj-N, and V-\_.

There are essentially two sets of MWEs, although the MWEs within them are mostly the same. The first set figured in the study relating to emotional valence; the second set figured in the study relating to arousal. The whole MWE ratings of valence and arousal were crowd-sourced using Amazon Mechanical Turk (AMT) (https://www.mturk.com/), The corresponding ratings of the constituent words stem from a list compiled by Warriner, Kuperman, and Brysbaert (2013). The list itself is available at: <http://crr.ugent.be/archives/1003>. See the main article for further details.

Associated with each set of MWEs is a smaller set of MWEs having both AMT ratings and ratings from Warriner et al. These matched, ‘overlapping’ ratings were used to assess the reliability of the new AMT ratings. In the main lists of MWEs these overlappers are given in red italics. Additionally, there are lists of regression residuals. The closer a residual is to zero, the more accurately MWE valence/arousal was predicted by the valence/arousal ratings of the mean of the constituent word ratings.

Important abbreviations used in the spreadsheet are: AMT = Amazon Mechanical Turk; C-word = Constituent word; Geo.mean = Geometrical mean; Harm.mean = Harmonic mean; Most valenced = The C-word rating that is the furthest from 5 (i.e., neutral) either toward 1 or toward 9; SD = The standard deviation of the individual AMT ratings obtained for a given MWE; WKB = Warriner et al. (2013); NA = not available.

NA was used in place of values (e.g., C-word ratings) that could not be found in the list of WKB. This abbreviation was chosen because the R functions used in the studies can handle datasets that include NA in place of a missing value. For instance, the appropriate calls in base R for calculating Spearman’s and Pearson’s correlationsbetween the variables x and y, when NAs are present, are: *cor(x, y, method = "s", use = "pairwise.complete.obs")* and *cor(x, y, method = "p", use = "pairwise.complete.obs").* The R functions of Wilcox (2012) that were used handle missing values even more automatically. For example, when Wilcox’s R functions are installed (https://dornsife.usc.edu/labs/rwilcox/software/), the call *corb(x,y, corfun = spear, nboot = 20000)* gives a bootstrap 95% confidence interval for Spearman’s correlation. The call for the median-based linear regression method that is mentioned would be: *tsreg(x, y)*, where x and y are the independent and the dependent variables, respectively.

**Reference**

Warriner, A., Kuperman, V., & Brysbaert, M. (2013). Norms of valence, arousal, and dominance for 13,915 English lemmas. *Behavior Research Methods, 45*, 1191–11207. https://doi.org/10.3758/s13428-012-0314-x