Predicting Lexical Norms Using a Word Association Corpus

Hendrik Vankrunkelsven (hendrik.vankrunkelsven@student.kuleuven.be)
Steven Verheyen (steven.verheyen@ppw.kuleuven.be)
Simon De Deyne (simon.dedeyne@ppw.kuleuven.be)
Gert Storms (gert.storms@ppw.kuleuven.be)
Faculty of Psychology and Educational Sciences, University of Leuven
Tiensestraat 102, B-3000 Leuven, Belgium

Abstract

Obtaining norm scores for subjective properties of words can be quite cumbersome as it requires a considerable investment proportional to the size of the word set. We present a method to predict norm scores for large word sets from a word association corpus. We use similarities between word pairs, derived from this corpus, to construct a semantic space. Starting from norm scores for a subset of the words, we retrieve the direction in the space that optimally reflects the norm data associated with the words. This direction is used to orthogonally project all the other words in the semantic space on, providing predictions of the words on the variable of interest. In this study, we predict valence, arousal, dominance, age of acquisition, and concreteness and show that the predictions correlate strongly with the judgments of human raters. Furthermore, we show that our predictions are superior to those derived using other methods.

Keywords: Similarity; MDS; Valence; Arousal; Dominance; Age of acquisition; Concreteness

Introduction

Lexical norm data are often asked for in psychological and linguistic research. Word properties like valence, arousal, dominance, concreteness, and age of acquisition (AoA), can guide the selection of experimental materials for manipulation or control of these crucial dimensions. Research on priming, lexical decision, and L2 learning, for example, often depend on the incorporation of these variables, and others (e.g., De Groot & Keijzer, 2000; Hinojosa, Carretié, Méndez-Bértolo, Míguez, & Pozo, 2009). Analysis of emotions also requires these norms in certain lines of research (e.g., Fossati et al., 2003).

Obtaining norm data can be quite a challenge as they generally require multiple human judgments for each of the words in what are generally large sets of words. In practice, this leads to a considerable investment of both the researcher's and participants' time. The investment can be alleviated, however, if reliable estimates of the ratings can be obtained through different means. In this paper, we propose and test a method for arriving at reliable proxies for a number of basic semantic dimensions on the basis of relatively small sets of words. Before describing the method in more detail, we briefly discuss the semantic dimensions we consider in our test of the approach.

Semantic dimensions

Arguably the three most important affective ratings are valence, dominance, and arousal, each of which is strongly rooted in semantic space (Osgood, Suci, & Tannenbaum, 1957). Valence, that is, the evaluation of pleasantness, is the affective variable most firmly present in semantic space (Osgood et al., 1957). Dominance, also labeled as potency, power, or control, is the second one. Arousal, an activity determinant, is the third. In three different analyses, using factor analysis, Osgood et. al. found that valence, dominance, and arousal explained a considerable proportion of the total variance of semantic meaning (valence 16% to 34%, dominance 7% to 8%, and arousal 5% to 6%). Moreover, this finding has been shown to hold for semantic spaces across cultures (Osgood, 1975).

Apart from the affective dimensions, we consider two other variables that have been shown to affect word processing and semantics: concreteness and age of acquisition (AoA). These variables are arguably the most essential non-affective variables based on subjective ratings (Brysbaert, Stevens, De Deyne, Voorspoels, & Storms, 2014). Concreteness refers to how well a word can be experienced by the senses. Easy perceivable words will lean towards the concrete pole of this dimension and unperceivable words will result in a rating towards the abstract pole. Furthermore, concreteness has been shown to be influential in memory and word processing, resulting in the adoption of concreteness in Paivio's dual-coding theory (Paivio, 1971, 2013) and the semantic theory of Vigliocco, Vinson, Lewis, and Garrett (2004).

AoA refers to the age at which a word is acquired during the language acquisition process. AoA has been shown to be an important variable in the organization of the mental lexicon, explaining about 5% of the variance in lexical decision times when other confounding variables such as word frequency are partialled out (Kuperman, Stadthagen-Gonzalez, & Brysbaert, 2012).

Extrapolating ratings for semantic dimensions

In light of the considerable investments required to arrive at ratings for a semantic dimension, researchers have recently attempted to predict lexical norm data from text corpora (Bestgen & Vincze, 2012; Recchia & Louwerse, 2014). In these studies, the co-occurrence of word pairs forms the basis from which the predictions are derived. Bestgen and

Vincze used the Touchstone Applied Science Associates (TASA) corpus, which consists of high-school text on a variety of academic topics. Recchia and Louwerse made use of the Google Web 1 T 5-gram corpus consisting of text from publicly accessible Web pages. These techniques typically yield promising correlations with subjective ratings, yet there is still room for improvement. For valence, for example, Bestgen and Vincze report a correlation of .71 and Recchia and Louwerse report one of .82. In this article, we present a similar method to extrapolate lexical norm data from a smaller set of subjective ratings making use of a large word association corpus instead of co-occurrence data.

Our method works as follows. First, a semantic space containing the words of interest is constructed using multidimensional scaling (MDS) with pairwise similarities between these words as input. The word similarities are not obtained from text corpora, but from a large-scale word association corpus. When a set of words with known values for a variable is included in the semantic space, it is possible to identify a direction in the semantic space that reflects this variable. This is done by property fitting (PROFIT) that is, regressing the norm scores on the coordinates of the corresponding words in the semantic space, allowing one to retrieve the direction in the geometric space that optimally matches the norm (Kruskal & Wish, 1978). This direction essentially is a line in the semantic space and can be used to project the rest of the words of interest on, providing an estimate for this variable for each of the words in the space.

In this paper we test the quality of the described method by comparing predicted norm scores with human data from two large norm datasets. Furthermore, to evaluate the robustness and cost-effectivity of the method, we vary the size of the observed word samples on the basis of which the norm scores are predicted for the remaining words. In the next section, we provide more detail on the sources of data.

Method

Lexical Norms

Norms for valence, arousal, dominance, and age of acquisition for Dutch words were obtained from data gathered by Moors et al. (2013). This dataset contains norms for 4,300 Dutch words that were collected from 224 university students, using a 7-point Likert scale. Each participant rated the entire set of words for one variable resulting in a total of 32 raters per word for AoA and 64 raters per word for the other variables.

Norms for concreteness were taken from Brysbaert et. al. (2014). This dataset has norm scores for approximately 30,000 Dutch words. Seventy-five university students rated one of five lists of 6,000 words, so every word was rated 15 times.

The reliability of the ratings of these variables was evaluated by applying the Spearman–Brown formula to the split-half correlations. All reliability indices were calculated on 10,000 different randomizations of the participants and the means of the different outcomes of these randomizations are the reliability coefficients we report here. The reliability coefficients for valence, arousal, dominance, and age of acquisition, from Moors et. al. (2013), are .99, .97, .96, and .97, respectively. The reliability coefficients of the concreteness ratings of Brysbaert et. al. (2014) for the five lists of 6,000 different words ranged from .91 to .93.

Word Similarities

Similarities between word pairs were obtained using the word association corpus reported in De Deyne, Navarro, and Storms (2013).

The collection of word associations started in 2003 and the most extensive version of the dataset is described in De Deyne et al. (2013). We used associations for a set of 12,566 cue words to obtain pairwise similarities between words. In line with our previous work, only responses that were part of the set of cues were retained, which transformed the cue x response matrix into a cue x cue matrix (De Deyne et. al., 2013). Starting from this square matrix with entries equal to the frequency with which the column word is given as a response to the row cue word, similarities were derived using the cosine measure (e.g., Landauer & Dumais, 1997) after applying a positive point-wise mutual information weighting scheme to avoid over-weighting high-frequency edges between words (e.g., De Deyne, Verheyen, & Storms, 2015). For the current study, similarities from 3,788 Dutch words were derived, that is, all the words that were both present in the word norms obtained by Moors et al. (2013), 4,300 in total, and in the word association corpus in the year 2012, that is 12,566 words. The resulting similarities were used as input for the construction of the semantic space. From these 3,788 words in the semantic space, 3,766 had an overlap with the concreteness norm scores.

Semantic Space

Nonmetric MDS (Kruskal & Wish, 1978) was employed to configure the semantic space. This technique constructs a multidimensional space where the resulting Euclidean distance between word pairs is as close as possible, inversely related, to the original similarities. Highly similar words are thus located close together in the obtained configuration and dissimilar words are further apart. We used High-Throughput MDS (HiT-MDS; Strickert, Teichmann, Sreenivasulu, & Seiffert, 2005) for its fast processing, and we obtained configurations in 2 to 30 dimensions (seeing that the predictions reach their maxima in 30 dimensions) to allow evaluation of an effect of dimensionality.

The obtained semantic space can be expected to encompass valence, arousal, and dominance, as Van Rensbergen, Storms, and De Deyne (in press) have shown that these variables strongly affect which concepts people regard as related. For instance, when presented with a cue-word of low arousal like ‘sleep’, people are more likely to give an association like ‘quiet’, which is also low in arousal/activity, than an association with high arousal like 'working'. Yet, it has not yet been established whether concreteness and AoA are represented in the semantic space.

Predicting the Norms

To predict norm scores for the variables of interest, a random subset of the words present in both the norm set and the association norms was used to find the corresponding direction in the semantic space that optimally predicts the norms of this subset of words. This was done using PROFIT where multiple linear regression is used with the norms in question as criterion and the coordinates of the words in the semantic space as predictors (Kruskal & Wish, 1978). The remaining words can then be orthogonally projected on this optimal direction and the resulting values serve as predicted norms.

As a quality measure of the prediction, the correlation between the predicted values and the corresponding human ratings was calculated for all available words, excluding those used to fit the optimal direction. For example, if 200 words were used to determine the optimal direction of the variable in the semantic space, the remaining 3,588 words (or 3,566 in the case of concreteness) served to calculate the correlation. This cross-validation technique was repeated for 200 random word samples. We report the mean of the correlation across these 200 random samples.

The sample size we primarily focus on is 200 words, yielding a ratio of .0557 (i.e., 200/3588) for valence, arousal, dominance, and AoA, and .0561 (i.e., 200/3566) for concreteness, between the word sample and the set for which scores were extrapolated. To gauge the effect of the sample size on the quality of the prediction, we used sample sizes of 50 to 500, with a step size of fifty.

Results

Before looking at the results of the analysis, it is important to appreciate that the theoretical maxima of the correlations between the empirically gathered and the predicted norms are not equal to 1.0, but have an upper limit that is not only related to how well the semantic space captures the predicted variables and the limitations of the method used (MDS) to construct this semantic space, but also to how reliable the human norms scores are. These maxima can be calculated by running a multiple linear regression with all of the data at hand. That is, by regressing all available norm scores on the coordinates of the corresponding words instead of using a sample of words. The root of R^2 (coefficient of determination) of this regression analysis defines this theoretical maximum, that is, the optimized correlation of the optimal dimension and the human ratings when all available data is used. Table 1 shows these maxima (max r) and R^2s for a 30 dimensional semantic space[1].

[1] We show these coefficients for a 30 dimensional space because this dimensionality provides good predictions as we will show later. The coefficients are typically smaller in lower dimensional spaces.

Aspects of the stimulus words that did not guide the participants in the word association task can, of course, not be detected in the constructed semantic space, as they have not determined the input similarities used for the MDS. Hence, the R^2 when predicting variables that quantify these aspects should be zero. The adjusted R^2s of the five criterion variables ranged from .52 to .82 (all p values < .001) in a solution with 30 underlying dimensions, illustrating their influence in the word association process, albeit some variables seem to have less of an influence on the association process than others and as a consequence, the semantic space derived from these associations does not fully capture these variables (e.g., AoA).

Table 1: Adjusted coefficients of determination (R^2) and correlation coefficients (max r) for a 30 dimensional solution. These values mark the theoretical maxima of what this method can achieve.

# Dimensions	30	
	R^2	max r
Valence	.82	.90
Arousal	.63	.79
Dominance	.64	.80
AoA	.52	.72
Concreteness	.70	.84

Evidently, the dimensionality of the semantic space and the sample size employed in the prediction of the norm scores have an impact on how well the predicted scores correlate with the norm scores as well: the higher the chosen dimensionality and the larger the sample size, the better the prediction (see Figures 1 and 2).

Figure 1 depicts the mean correlations of the predictions of 200 random samples of size 200 as a function of dimensionality. The variability in the correlations over the 200 different samples is shown as 90% highest density intervals (HDI) with vertical bars. The HDI's for the different variables indicate that the spread of these correlations is quite small, thus making the predictions from random samples fairly consistent.

As can be seen in Figure 1, for variables with an R^2 higher than .60 (all except AoA), adding dimensions beyond 17 does not benefit the quality of the prediction substantially. The prediction of AoA on the other hand does benefit from adding more dimensions and does not seem to converge as smoothly to its asymptote (The horizontal lines, next to dimension 30, give the theoretical maxima the correlations can reach for each variable. See Table 1).

In the rest of this paper we present results based on a semantic space of 30 dimensions as the predictions are more valid in higher dimensional spaces. When the quality of the predictions cannot be assessed through comparison with existing norm scores, we propose running MDS multiple times using a different amount of dimensions (preferably over 20) and then choosing the dimensionality where the adjusted R^2, from the linear regression used to determine a direction in the semantic space, converges to a maximum. However, when this R^2 is small, the variable under consideration is not captured by the semantic space, therefore, the predictions will not be trustworthy.

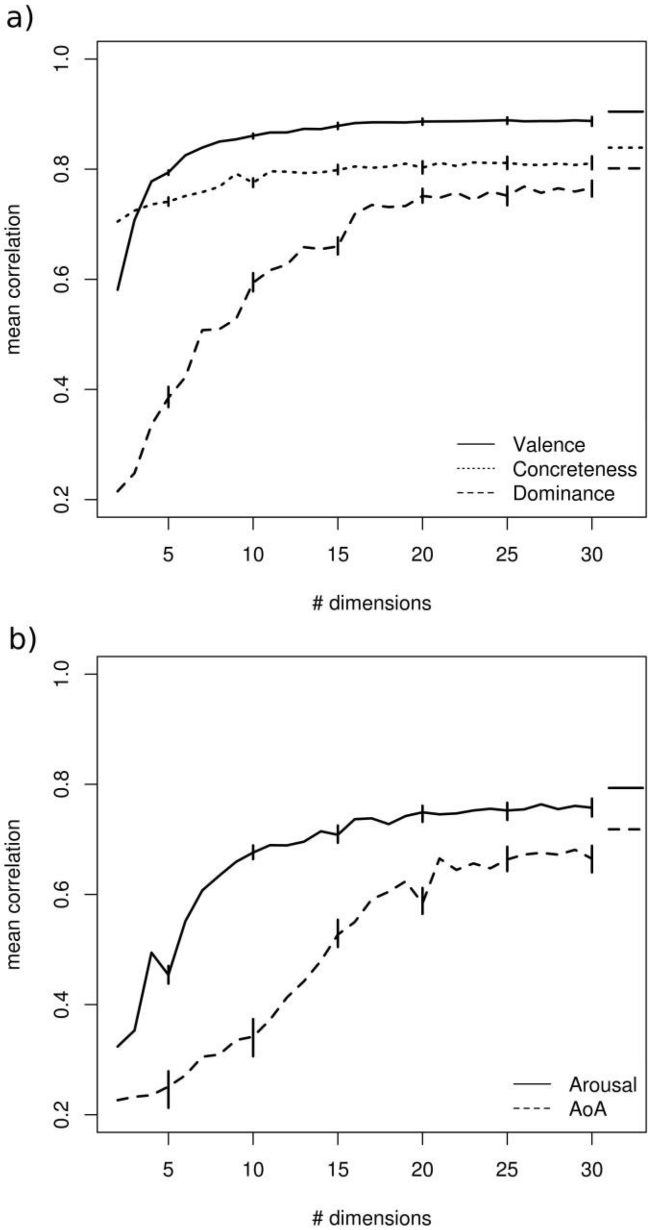

Figure 1: Mean correlations of the predictions of 200 random samples of size 200 as a function of dimensionality for valence, concreteness, and dominance (a) and arousal and AoA (b). The horizontal lines, next to dimension 30, give the maxima the correlations can reach for each variable in a 30 dimensional space. The vertical lines give the 90% highest density intervals from the sampled distribution.

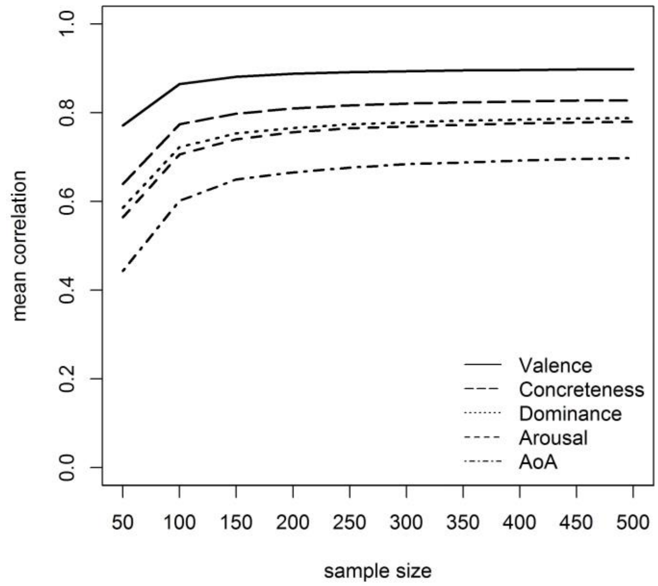

Figure 2: Mean correlations of the predictions of 200 random samples of size 50 to 500 with steps of 50 in a 30 dimensional space.

In Figure 2 the effect of sample size on the prediction is illustrated for a semantic space with 30 dimensions. Clearly, the sample size used to predict the norms can be relatively small. Regardless of the norm variable that is predicted, the quality of the prediction improves a lot when the sample size increases from 50 to 100 words, but gains little beyond sample sizes of 200, signifying the limited amount of norm score data needed when employing this method.

Correlations

Table 2 lists the mean *r* between predicted scores and norm scores of 200 random samples with sample size 200, the means of the adjusted R^2s (not of the full dataset but of the 200 random samples), and the standard deviations of these adjusted R^2s from the samples, for a 30 dimensional MDS space, alongside predictions using text corpora from other authors.

Valence clearly has the highest prediction quality. It has a mean correlation of .89. Regardless of the method used for predicting norm scores, the upper limit of this correlation is confined to the reliability of the norm scores one is correlating them with. For valence the split-half reliability of the full dataset of Moors et al. (2013) is .99. The mean prediction of arousal is .76. The split-half reliability for arousal from the data of Moors et. al. is .97. Dominance reaches a mean correlation of .77 using our method. Moors et al. obtain a reliability of .96 for this variable. For AoA, the obtained correlation is .67, while AoA obtained by Moors et al. has a reliability of .97. Finally, concreteness measured by Brysbaert et al. (2014) has a split-half reliability of about .93 and a correlation of .81 is reported here. In all cases, these predictions correlated more with human norms than comparable methods that use text corpora (see conclusion).

Furthermore, instead of a semantic space with 3,788 words used for prediction in the aforementioned results, we also used all the words both present in the concreteness norm data and the association corpus (11,547 words) to construct a 30 dimensional semantic space. The mean correlation for concreteness using this space, using samples of 200 words to predict the remaining 11,347 words, was .80. This prediction is on par with the correlation of .81 we obtained in predicting concreteness for 3566 words.

Table 2: Mean correlations (mean *r*) between the predictions and human ratings based on 200 random samples of sample size 200, the adjusted mean R^2s (mean R^2) used to obtain the direction in the semantic space for the 200 samples, and the standard deviation (*SD* R^2) of these adjusted R^2s using our method. Correlations of predicted norms with the ANEW norms and the Warriner norms from Bestgen and Vincze (2012; B&V), and Recchia and Louwerse (2014; R&L). (Val = Valence, Aro = Arousal, Dom = Dominance, AoA = Age of acquisition, Con = Concreteness)

Method	Measure	Val	Aro	Dom	AoA	Con
Our	mean *r*	.89	.76	.77	.67	.81
	mean R^2	.81	.63	.64	.51	.70
	SD R^2	.02	.04	.04	.05	.04
B&V	*r* ANEW	.71	.56	.60	-	.79
R&L	*r* ANEW	.80	.62	.66	-	-
	r Warr.	.82	.64	.72	-	-

Conclusion

We presented a method to estimate norm scores for variables that are incorporated in a semantic space derived from word association data. Using a relatively small set of words for which human norm scores are known, we derived an optimal direction in this space and by projecting the remaining words in the space on this direction, we obtained estimates.

The extrapolation method presented in this article is shown to have a good validity for semantic variables that are well embedded in the semantic space. The quality of the estimates differs as a function of how well the semantic space captures the predicted variables. For variables that are well captured in the space, like valence, the obtained predictions reach very high correlations (.89) with human ratings, especially when considering that these predictions are also attenuated by the not-perfect reliability of the norms used to find the corresponding direction in space. For variables like AoA, the predictions are clearly of lower quality, but are stable from 21 dimensional solutions, and from a sample size of 200, onwards.

Other techniques (see Table 2) to predict word norm scores have been described in the literature (Bestgen & Vincze, 2012; Recchia & Louwerse, 2014). These authors extracted a semantic space from English text corpora and predicted norms using the *k* nearest neighbors method. Using different English norm datasets (Bradley & Lang, 1999; Warriner, Kuperman, & Brysbaert, 2013), Bestgen and Vincze reported correlations of .71, .56, and .60, for valence, arousal, and dominance, respectively, and Recchia and Louwerse reported correlations of .80 and .82 for valence, .62 and .64 for arousal, and .66 and .72 for dominance. The method described in the current article exceeds these alternative predictions, reaching correlations of about .89, .76, and .77 for valence, arousal, and dominance, respectively. Bestgen and Vincze also report predictions for concreteness that are on par with the predictions in this article: .79 (reported by Bestgen & Vincze) vs. .81 (reported here).

The corpora Bestgen and Vincze (2012) and Recchia and Louwerse (2014) used are different from the association corpus we used here. First, they are English corpora and the one we used is Dutch. A word association corpus in English is available (http://www.smallworldofwords.com), with currently over one million association responses. A systematic comparison of the norm score predictions using the English and the Dutch word association corpus is planned. Second, the text corpora used by Bestgen and Vincze, and Recchia and Louwerse are a lot bigger than the corpus we used, making it possible to predict more words. However, the Dutch association corpus already consists of over 16,000 words and is constantly expanding, and similar studies in different languages are currently on their way. It will therefore be possible to predict norm scores for an even larger set of stimulus words as the word association corpus grows. Third, De Deyne, Verheyen, and Storms (2015) demonstrated that making use of associations to capture human judgments of similarity is superior to using text to capture similarity. The information captured in association corpora seems to consist of a wider array of semantic and lexical properties, enabling the prediction of even very weak semantic relations (De Deyne, Navarro, Perfors, & Storms, 2012). Fourth, unlike text corpora, it is straightforward to use word association corpora to tailor norms to specific populations (men vs. women, young vs. old) when required (De Deyne & Storms, 2007). It suffices to employ only the associations from members of these populations to build a tailored semantic space. Aside from the different corpora used, the human norm scores used to compare the predicted norms with, were also different.

The reported estimates can still be improved upon. In this paper we have shown how to extrapolate norm scores from a small sample of human ratings. But, larger datasets of human ratings are available and therefore it is possible to include these ratings to find a more reliable direction in semantic space used for prediction. Thus, when combining the word-similarities of the desired set and these of a large set of reliable norms, estimates can reach correlations that are almost the same as the theoretical maxima. For instance, when using 3588 of the 3788 words from Moors et. al. (2013) to obtain this direction in semantic space, the remaining 200 words can be predicted with an accuracy of .90, .79, .80, and .72 for valence, arousal, dominance, and AoA, respectively. Predictions for concreteness using 3566 words from the 3766, reach a correlation of .83[2].

Acknowledgments

We would like to thank Bram Van Rensbergen and Wouter Voorspoels, two fellow members of the ConCat (concepts and categories) research lab at the University of Leuven, for contributing valuable ideas. Steven Verheyen is a postdoctoral fellow at the Research Foundation – Flanders. This study was partly sponsored by Grant G.0436.13 of the F.W.O.-Flanders, awarded to Gert Storms, and by an ARC grant, DE140101749, awarded to Simon De Deyne.

References

Bestgen, Y., & Vincze, N. (2012). Checking and bootstrapping lexical norms by means of word similarity indexes. *Behavior Research Methods, 44*, 998–1006.

Bradley, M. M., & Lang P. J. (1999). *Affective norms for English words (ANEW): Stimuli, instruction manual and affective ratings.* (Tech. Rep. No. C-1). Gainesville, FL: Center for Research in Psychophysiology, University of Florida.

Brysbaert, M., Stevens, M., De Deyne, S., Voorspoels, W., & Storms, G. (2014). Norms of age of acquisition and concreteness for 30,000 Dutch words. *Acta Psychologica, 150*, 80–84.

De Deyne, S., Navarro, D. J., Perfors, A., & Storms, G. (2012). Strong structure in weak semantic similarity: A graph based account. In *Proceedings of the 34th annual conference of the cognitive science society* (pp. 1464–1469). Austin, TX: Cognitive Science Society.

De Deyne, S., & Navarro, D.J., & Storms, G. (2013). Better explanations of lexical and semantic cognition using networks derived from continued rather than single word associations. *Behavior Research Methods, 45*, 480-498.

De Deyne, S., & Storms, G. (2007). Age-of-acquisition differences in young and older adults affect latencies in lexical decision and semantic categorization. *Acta Psychologica, 124*, 274-295.

De Deyne, S., & Storms, G. (2008). Word associations: Norms for 1,424 Dutch words in a continuous task. *Behavior Research Methods, 40*, 198–205.

De Deyne, S., Verheyen, S., & Storms, G. (2015). The role of corpus size and syntax in deriving lexico-semantic representations for a wide range of concepts. *Quarterly Journal of Experimental Psychology.* doi:10.1080/17470218.2014.994098

De Groot, A. M. B., & Keijzer, R. (2000). What is hard to learn is easy to forget: The roles of word concreteness, cognate status, and word Frequency in foreign-language vocabulary learning and forgetting. *Language Learning, 50*, 1–56.

Fossati, P., Hevenor, S. J., Graham, S. J., Grady, C., Keightley, M. L., Craik, F., & Mayberg, H. (2003). In search of the emotional self: An fMRI study using positive and negative emotional words. *American Journal of Psychiatry, 160*, 1938–1945.

Hinojosa, J. A., Carretié, L., Méndez-Bértolo, C., Míguez, A., & Pozo, M. A. (2009). Arousal contributions to affective priming: Electrophysiological correlates. *Emotion, 9*, 164–171.

Kruskal, J. B., & Wish, M. (1978). *Multidimensional scaling*. Beverly Hills, CA: Sage.

Kuperman, V., Stadthagen-Gonzalez, H., & Brysbaert, M. (2012). Age-of-acquisition ratings for 30,000 English words. *Behavior Research Methods, 44*, 978–990.

Landauer, T. K., & Dumais, S. T. (1997). A solution to Plato's problem: The latent semantic analysis theory of acquisition, induction, and representation of knowledge. *Psychological Review, 104*, 211–240.

Moors, A., De Houwer, J., Hermans, D., Wanmaker, S., van Schie, K., Van Harmelen, A.-L., … Brysbaert, M. (2013). Norms of valence, arousal, dominance, and age of acquisition for 4,300 Dutch words. *Behavior Research Methods, 45*, 169–177.

Osgood, C. E. (1975). *Cross-cultural universals of affective meaning*. Urbana, IL: University of Illinois Press.

Osgood, C. E., Suci, G. J., & Tannenbaum, P. H. (1957). *The measurement of meaning*. Urbana, IL: University of Illinois Press.

Paivio, A. (1971). *Imagery and verbal processes*. Austin, TX: Holt, Rinehart and Winston.

Paivio, A. (2013). Dual coding theory, word abstractness, and emotion: A critical review of Kousta et al. (2011). *Journal of Experimental Psychology: General, 142*, 282–287.

Recchia, G., & Louwerse, M. M. (2014). Reproducing affective norms with lexical co-occurrence statistics: Predicting valence, arousal, and dominance. *The Quarterly Journal of Experimental Psychology, 68*, 1–15.

Strickert, M., Teichmann, S., Sreenivasulu, N., & Seiffert, U. (2005). High-throughput multi-dimensional scaling (HiT-MDS) for cDNA-array expression data. In W. Duch, J. Kacprzyk, E. Oja, & S. Zadrożny (Eds.), *Artificial neural networks: Biological Inspirations.* Berlin Heidelberg: Springer.

Van Rensbergen, B., Storms, G., & De Deyne, S. (in press). Examining assortativity in the mental lexicon: Evidence from word association data. *Psychonomic Bulletin & Review.* doi: 10.3758/s13423-015-0832-5

Vigliocco, G., Vinson, D. P., Lewis, W., & Garrett, M. F. (2004). Representing the meanings of object and action words: The featural and unitary semantic space hypothesis. *Cognitive Psychology, 48*, 422–488.

Warriner, A. B., Kuperman, V., & Brysbaert, M. (2013). Norms of valence, arousal, and dominance for 13,915 English lemmas. *Behavior Research Methods, 45*, 1191–1207.

[2] These correlations are average correlations across 200 random samples of the words to find the optimal direction.