Towards a Cognitively Realistic Representation of Word Associations

Ivana Kajić (ivana.kajic@plymouth.ac.uk)[1,2]
Jan Gosmann (jgosmann@uwaterloo.ca)[1]
Terrence C. Stewart (tcstewar@uwaterloo.ca)[1]
Thomas Wennekers (thomas.wennekers@plymouth.ac.uk)[2]
Chris Eliasmith (celiasmith@uwaterloo.ca)[1]
[1] Centre for Theoretical Neuroscience, University of Waterloo
Waterloo, ON, Canada N2L 3G1
[2] School of Computing, Plymouth University
Plymouth, Drake Circus, United Kingdom PL4 8AA

Abstract

The ability to associate words is an important cognitive skill. In this study we investigate different methods for representing word associations in the brain, using the Remote Associates Test (RAT) as a task. We explore representations derived from free association norms and statistical n-gram data. Although n-gram representations yield better performance on the test, a closer match with the human performance is obtained with representations derived from free associations. We propose that word association strengths derived from free associations play an important role in the process of RAT solving. Furthermore, we show that this model can be implemented in spiking neurons, and estimate the number of biologically realistic neurons that would suffice for an accurate representation.

Keywords: semantic spaces; vector representations; spiking neurons; insight; remote associates test

Introduction

Creating word associations is an important skill for the development of many cognitive abilities. Language acquisition is highly dependent on the ability to associate words (Elman et al., 1997), and the associative organisation of children's language is known to facilitate learning of new words and syntax (Brown & Berko, 1960; Hills, 2013). Furthermore, word associations have been shown to play a role in analogical problem solving important for inference and concept attainment (Powell & Vega, 1971).

Despite their importance in a variety of cognitive tasks, it is not clear how associations are represented in the brain. This question is of interest to researchers in cognitive and computer science. From a cognitive science perspective, understanding word representations has been relevant for the modelling and explaining of psycholinguistic phenomena (Landauer & Dumais, 1997; Jones & Mewhort, 2007; Steyvers, Shiffrin, & Nelson, 2004). In computer science, machine learning has been concerned with optimal word representations for automated natural language processing and text comprehension.

In this study we investigate biologically plausible representations of word associations. To this end, we analyse two sources of information about word associations and different forms of encoding of the associations. We compare these methods on predicting human performance on the Remote Associates Test (RAT). In particular, we note that some methods of representation may be better for solving this task, but those methods may not do as good a job at predicting human performance. Finally, we take the representation method that is closest to human performance and implement it using spiking neurons. Not only does this demonstrate that this algorithm could be implemented biologically, but it also allows us to determine how many neurons would be needed to represent these associations, given realistic biological constraints.

Remote Associates Test

The RAT was conceived to study the ability of an individual to form new associations among seemingly unrelated words (Mednick, 1962). The test consists of a list of word problems and each problem contains three cue words (e.g., *call*, *pay*, *line*). The task is to find a word (*phone*) associated with all three cue words within a time limit of up to several seconds. The words can form a word phrase (phone line), or a compound word (payphone). Individuals scoring higher on the test are assumed to more easily create uncommon and less stereotypical associations between pairs of words. The RAT has been used in psychology and cognitive neuroscience to study creative thinking and insight. Because the RAT problems differ in difficulty, they give us information about which associations are common and therefore easier to come up with. This allows us to infer which ways of representing associations are likely used in the brain as they should reproduce the same patterns of easy and hard problems.

For comparison to human data, this work uses data set from the experimental condition where subjects were given only a few seconds to solve a problem. This is meant to address situations where people are solving RAT based on insight, rather than explicitly searching for solution. Longer time periods would allow analytical solving, rather than relying on unconscious information such as word association, and lead to higher scores on the RAT (Bowden & Jung-Beeman, 2003; Kounious & Beeman, 2014).

Related Work

A number of studies have investigated the influence of word representations in in a wide range of semantic memory tasks (Landauer & Dumais, 1997; Steyvers et al., 2004; Jones & Mewhort, 2007). Common to all approaches is the representation of words as vectors, whose relationships can be quantified using linear algebra methods. Latent Semantic

Analysis (LSA) evaluates word occurrences in large corpora of text and derives vector representations for each word. The words similar in meaning will have similar vector representations. LSA has been successfully applied to explain a variety of psycholinguistic phenomena (Landauer & Dumais, 1997; Deerwester, Dumais, Furnas, Landauer, & Harshman, 1990). Instead of using large corpora of text to derive the semantic spaces, another approach is to use free association norms (FAN; Nelson, McEvoy, & Schreiber, 2004). The association strengths between word pairs have been derived experimentally, by asking each participant to provide the first word which comes to their mind given a word cue. By applying dimensionality reduction techniques on this data, the word association space (WAS) was created and used to predict the performance on semantic memory tasks such as recognition and recall (Steyvers et al., 2004). Corpus-based approaches have been shown to solve the RAT with solution rates higher than humans (Toivonen, Gross, Toivanen, & Valitutti, 2013; Klein & Badia, 2015). However, very few studies have investigated models which match the performance on the RAT with the human performance (but see, Kajić & Wennekers, 2015; Bourgin, Abbott, Griffiths, Smith, & Vul, 2014; Gupta, Jang, Mednick, & Huber, 2012). In this study we analyse what kind of biologically plausible representations yield a performance comparable to human performance on this test.

Representation of associations

We use two datasets to construct the word representations. The first dataset are the free association norms (Nelson et al., 2004), containing association strengths for over 5000 word cue-target pairs. The strengths between the words are organized in an asymmetric association matrix FAN_{asym}, with rows representing cues and columns representing targets used in the free association experiment. The asymmetry is a result of non-reciprocal association strengths. For example, given the cue *left*, 94% subjects respond with *right*, however, given the cue *right*, 41% subjects respond with *left* and 39% subjects respond with *wrong*. Formally, this is a difference in the forward (cue to target) and the backward (target to cue) association strength. In addition to the asymmetric matrix, the symmetric matrix FAN_{sym} is created by adding the asymmetric matrix and its transpose.

The second form of association information is derived from the Google Books Ngram Viewer dataset (version 2 from July 2012; Michel et al., 2011). An n-gram is a sequence of n words, and this dataset provides occurrence frequencies of n-grams across over 5 million books published up to 2008. The set of words used in this study is restricted to the same words that have been used in the FAN data. Furthermore, we only used n-gram frequencies from 2008. For every combination of two words w_1 and w_2 the corresponding entry in the matrix NG_{asym} was set to the sum of occurrences of the 2-gram (w_1, w_2) and the 1-gram $w_1 w_2$ in the corpus. Each row of the matrix was then normalized to sum to one. The symmetric matrix $NG_{\text{sym}}$is computed in the same way as the FAN_{sym}matrix. In the rest of the analysis we will merely use NG_{sym} which includes the backward strength between word co-occurrences. This is necessary to solve the problems where only the second part of the compound word is given as one of the three cues (e.g., *board* for *blackboard*). Even though the NG matrices give co-occurrence counts, we will use the terms *association matrix* and *association strength* as they are used in the same manner as the FAN association matrix.

There are two commonly used approaches to represent word associations. First, we can directly use the association matrix. That is, we represent a word as a localist vector (all zeros except for a single one for the word itself), and then, to perform the association we multiply the word by the association matrix. The non-zero entries in the resulting vector represent the word associates. Alternatively, we can embed the associates in a vector space. That is, instead of representing the full association matrix, we compress that matrix into a lower-dimensional space. In certain cases this approach can adjust the similarity space between the words to uncover latent structure among the associations. For example, this is the basis for Latent Semantic Analysis (Deerwester et al., 1990), where similar words are made more similar and less similar words less similar. In particular, we use singular value decomposition (SVD) to take the 5018-dimensional localist word representation and compress it into an D-dimensional distributed representation (where D is varied between 128 and 4096).

Preliminary evaluation

To determine which representational approach gives the best performance on the task, we use the problem set from Bowden and Jung-Beeman (2003). Out of 144 RAT problems, we used the 117 problems for which the cues and the target exist in the set of free association norms. We took the sum of the vector representations of each cue word, and multiplied it by the association matrix. The resulting vector was compared to the vectors for all of the possible response words. In the ideal case, the correct solution word would be the most similar to this output value. However, we also determined if the solution word was in the top 2, 3, 5, and 10 most similar words, as reported in Table 1.

The results indicate that the solution appears as the top-ranked word more often for Google n-grams (NG_{sym}) than for association norms (11 solutions in the first position versus

Table 1: Target positions for 117 RAT problems

	Within top				
Association matrix	1	2	3	5	10
FAN_{asym}	5	12	16	31	49
FAN_{sym}	4	5	6	14	36
NG_{sym}	11	15	16	22	35

5 and 4 solutions for symmetric and asymmetric matrices). However, if we allow for the solution word to be in the top 5 or 10, then the FAN_{asym} association matrix performs best.

Figure 1 also includes the results from applying SVD to the association matrices. Contrary to expectation, SVD does not improve the performance on the RAT test except in a few specific circumstances. In particular, if we need the solution word to be in the top 3 words and we are using the Google n-grams (NG_{sym}), then using 512-dimensional SVD provides a slight improvement over the full asymmetric FAN matrix.

In majority of cases, the statistical n-gram data performs better than the free association norms. However, this only shows increased performance on the task, not whether the n-gram approach performs similarly to people on this task. To address this question, we compare the two approaches with the human performance.

Matching human performance

Instead of analysing which method gets the most correct solutions on the RAT, we now explore which method yields the results most similar to human results. That is, we are interested to find which method is better in solving problems that humans find easy, and worse in solving problems humans find hard. To do so, we predict a probability of producing the correct solution within a 2 s time limit. We use the same set of problems as in the previous section, and match to the percentage of people solving each problem (Bowden & Jung-Beeman, 2003).

Let $s(w,v)$ be the associative strength from word w to v. Given the three cues c_k with $k = 1,2,3$ each word w_i in the vocabulary is activated according to

$$a(w_i) = \sum_{k=1}^{3} \alpha_k \cdot s(c_k, w_i) \tag{1}$$

where α_k are free parameters intended to model the effect that subjects might differently prioritize the problem cues. We set $a(c_k) = 0$ for the cues to prevent them from appearing high in the results. Moreover, we fix $\alpha_1 = 1.0$ as a scaling of all α_k with a constant will produce the same predictions.

Given that w_s is the solution word, we calculate the predicted probability for producing the correct answer as

$$P = \beta \cdot \frac{a(w_s)}{\sum_i a(w_i)} \tag{2}$$

with β being another free parameter. Note, that we are not calculating the probability of each individual word being given as answer, but the probability of producing the correct vs. the wrong response. Because of that, β is not fixed to one, but should be chosen such that $P \leq 1$.

We did curve fits to the data from Bowden and Jung-Beeman (2003) by minimizing the root mean square error between the proportion of participants solving the problem within the time limit and our predicted solving probability. For the curve fits we used the association strengths from the original FAN_{asym}, FAN_{sym}, and NG_{sym} matrices. In addition, we used the 768-dimensional NG^{768}_{sym} matrix which gave improved performance in Figure 1.

The resulting parameter values are given in Table 2. Representations derived from free norms yield a better fit on this data set ($r^2 = 0.58$) compared to the n-gram data ($r^2 = 0.30$). There was no difference between the asymmetric and symmetric FAN matrices. Interestingly, the second cue gets consistently a higher weight. We speculate that this is caused by this cue appearing in the center of the screen with the other cues above and below. For n-gram fits the parameters α_2 and β are large, but because of the low r^2-value, these values cannot be seen as meaningful. For visual inspection, we have plotted the model fits using free association norms and Google n-grams in Figure 2. Further error analysis revealed that the Google n-grams underestimate the solution probabilities of easy items (more than 32% solved by humans) while at the same time predicting a non-zero probability for items unsolved by humans.

All solutions in the data set used are based on compound words which explains that n-gram data can solve more and harder RAT problems. This also means that the insight process in RAT solving could be based on such co-occurrence information. But the results provide evidence that this is not the case and that the insight process is based on associations closer to the associations produced in an unconstrained free association task. These kinds of associations are likely to be based on additional semantic information not available to purely statistical approaches.

Biological plausibility

While the previous sections argue that people use word association data of the form seen in the free association norms to perform the RAT task, there is the separate question of whether such an operation can be implemented in the brain. Can neurons in the brain precisely implement the mathematical matrix operations described above? How many neurons would be needed to implement it accurately enough?

To determine this, we implemented the above algorithm using two groups of spiking Leaky Integrate-and-Fire (LIF) neurons, with synaptic connections from the first group to the second group. The first group represents the input (the sum of the vectors from the three cue words), and the second group represents the output (the result after multiplying by the association matrix).

To allow a group of neurons to represent a vector (which, in

Table 2: Model fits and best fitting parameters

Association matrix	α_1	α_2	α_3	β	r^2
FAN_{asym}	1.0	2.06	1.20	1.13	0.58
FAN_{sym}	1.0	2.50	1.63	2.86	0.58
NG_{sym}	1.0	13.45	1.25	3.55	0.30
NG^{768}_{sym}	1.0	11.88	1.00	8.23	0.22

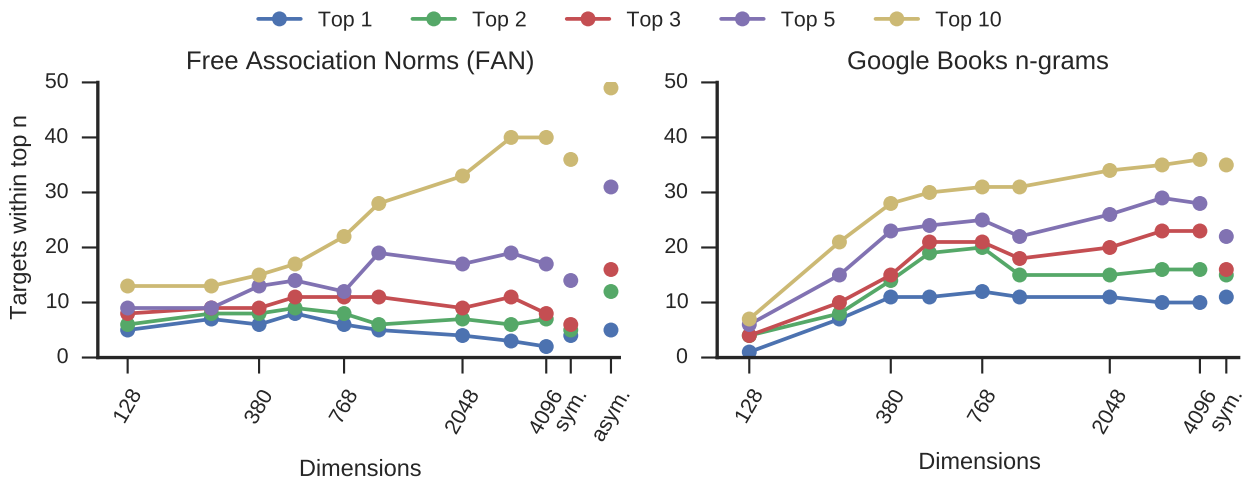

Figure 1: The number of correct solutions within the top n most similar words over 117 problems. The left plot shows the results with representations based on free association norms. The right plot is based on Google n-gram data. The isolated points at the end of the x-axis in both graphs represent the original symmetric and (for FAN) asymmetric matrices. All other data points are computed by applying the SVD to the matrices.

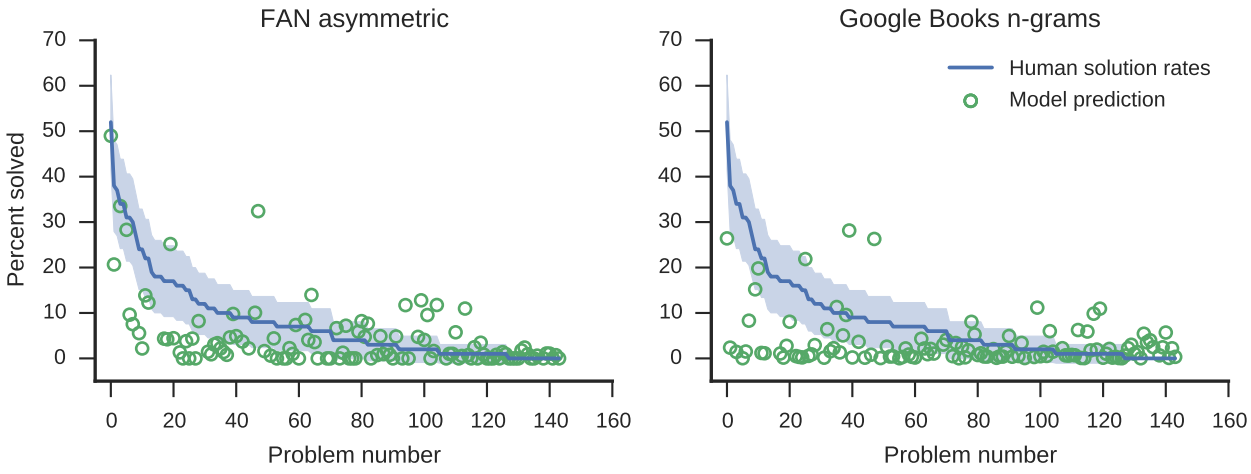

Figure 2: Model fits to human data (blue line with shaded 95 % confidence interval). Human data are percentages of participants solving a RAT problem in 2 s and they have been sorted in descending order, so that problems solved by more participants have lower problem indices. Every green circle represents the probability of producing the correct solution to a RAT problem predicted by the model.

turn, represents a word), each neuron in the group has a randomly chosen preferred vector. This is the vector for which this neuron will fire maximally. For other vectors, the neuron will fire less frequently. This is a generalization of the standard idea of neurons having a preferred stimulus or representation, as seen in motor cortex, visual cortex, and throughout the brain. Importantly, since each neuron's preferred vector is randomly chosen, the neurons will provide a distributed representation, even if the vector representation is localist. For example, if the represented vector is [0, 1, 0, 0], then one neuron might have a preferred vector of [0.1, 0.3, 0.8, 0.5] and so it would fire slowly (the similarity between the two vectors is 0.3, as measure by the dot product), while another neuron might have a prefered vector of [0.2, 0.9, 0.1, 0.4], so it would fire quite frequently (dot product of 0.9). We have previously shown that such a representation is extremely robust to neural damage and consistent with observed patterns of neural activity (Stewart, Bekolay, & Eliasmith, 2011).

Given this approach to representation, we need to connect the first group of neurons to the second group of neurons in such a way that if we cause the first group of neurons to fire as they should when representing a particular cue word vector x, then this should cause the second group of neurons to fire with the pattern for the vector that is the result of multiplying x by the association matrix. We do this by setting the synaptic weights between the first and second groups. Many techniques could be used to perform this task (including standard backpropagation learning rules), but here we simply treat it as a least-squares optimization problem and directly solve for the best set of connection weights for this task. This overall approach is known as the Neural Engineering Framework (NEF; Eliasmith & Anderson, 2003).

To test the model we used three RAT problems of easy, intermediate, and hard difficulty, as shown in Figure 3B. We estimate the output similarities from the spiking output with the methods of the NEF and compare it to the analytical result. The accuracy of the neural representation increases as the number of neurons increases. The root mean square error with the analytical result, relative to the word most similar to the cues, is in the range from 4.5 % to 3.0 % depending on the number of neurons ranging from 100 360 up to ten times as many neurons. Thus, we can approximate the model equations with biologically realistic spiking neurons with minor deviation.

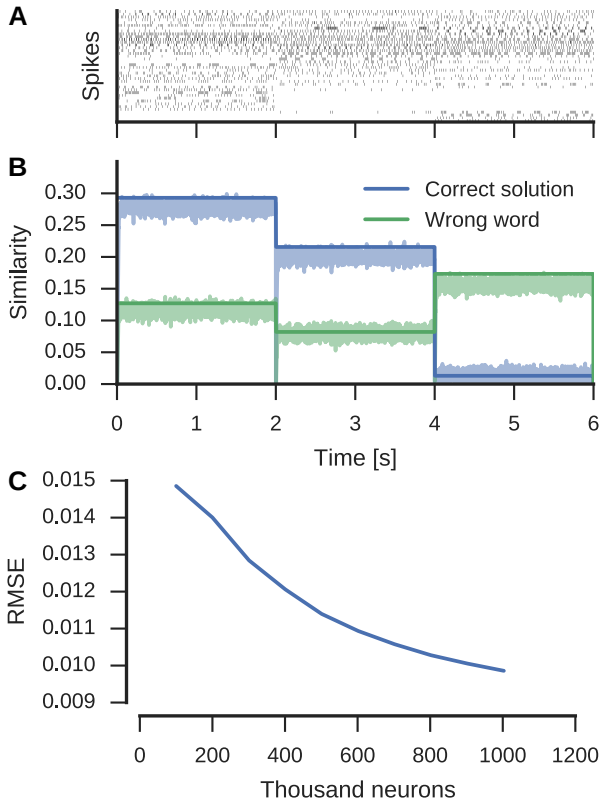

Figure 3: Example run of the neural network model for three RAT problems of easy, intermediate and hard difficulty. (**A**) Spike patterns of a subset of the neurons in the *Solution* ensemble. (**B**) Similarity of the representation in the *Solution* ensemble with the correct solution (blue) and most similar wrong word (green). The solid lines give the analytical result, whereas the semi-transparent lines give the network output. (**C**) Root mean square error between neural network output and analytical calculation as we change the number of neurons.

Discussion

In this study we have done a computational analysis of two different sources of word associations and described how well they predict human performance on the RAT. We have shown that statistical language data like n-grams allow the highest solution rates on this task, consistent with the previous work (Klein & Badia, 2015; Toivonen et al., 2013). However, further analysis revealed that the better prediction of the human performance is obtained with the free association norms.

First, we discovered structural differences between the n-gram data set and association norms by applying dimension reduction with SVD. Previous studies have shown that the dimensionality reduction on association norms can be used to accurately predict human responses on certain episodic memory tasks such as recognition memory and cued word recall (Steyvers et al., 2004). Our analysis provides evidence that dimensionality reduction does not improve performance on the RAT. Moreover, it impairs the performance when the target is among the ten most similar words to the cue words. This indicates that for some RAT problems the important associations are contained in links which are not present in a low-dimensional representation. This is reminiscent of the finding that direct association strengths are the best predictor of intrusion rates in free recall (Steyvers et al., 2004). Whether there is a connection between the associative mechanisms in the RAT and free recall with intrusion remains to be explored in the future work. The dimensionality reduction on n-grams revealed a considerable amount of redundant infor-

mation: the original 5018 dimensional word vectors can be reduced to at least 768 dimensions without large differences in the results. Moreover, the SVD can even lead to improvement when looking at targets which appear within the three most similar words to the cues. Second, the modelling analysis showed that n-gram data, yielding best scores on the RAT, is a worse predictor of human performance.

The FAN data model was a better fit to human solution probabilities in the RAT. As expected, the model was not able to solve any problems for which there was no association between the cues and the target in the association norms indicating that free norms might not be the only source of information. However, for the other problems we have demonstrated that free associations play an important role in the insight process.

Finally, we demonstrated the biological plausibility of this approach in a spiking neural model of the insight solution process of the RAT. The model shows the expected behaviour and is more likely to produce the correct solution for easy RAT problems. In the future, we plan to match the model more rigorously to human data and extend it with recurrent processing to explore a variety of mechanisms and representations involved in the memory search.

Notes

The model and data analysis source code are available at `https://github.com/ctn-archive/kajic-cogsci2016`.

Acknowledgments

This work has been supported by the Marie Curie Initial Training Network FP7-PEOPLE-2013-ITN (CogNovo, grant number 604764), the Canada Research Chairs program, the NSERC Discovery grant 261453, Air Force Office of Scientific Research grant FA8655–13–1–3084, CFI and OIT. This work made use of SHARCNET and Compute Canada computer resources.

References

Bourgin, D. D., Abbott, J. T., Griffiths, T. L., Smith, K. A., & Vul, E. (2014). Empirical Evidence for Markov Chain Monte Carlo in Memory Search. In *Annual Conference of the Cognitive Science Society.*

Bowden, E. M., & Jung-Beeman, M. (2003). Normative data for 144 compound remote associate problems. *Behavior Research Methods, Instruments, & Computers: A Journal of the Psychonomic Society, Inc*, *35*(4), 634–639.

Brown, R., & Berko, J. (1960). Word association and the acquisition of grammar. *Child Development*, *31*(1), 1–14.

Deerwester, S., Dumais, S. T., Furnas, G. W., Landauer, T. K., & Harshman, R. (1990). Indexing by latent semantic analysis. *Journal of the American Society for Information Science*, *41*, 391–407.

Eliasmith, C., & Anderson, C. H. (2003). *Neural engineering: computation, representation, and dynamics in neurobiological systems*. Cambridge, MA: MIT Press.

Elman, J. L., Bates, E. A., Johnson, M. H., Karmiloff-Smith, A., Parisi, D., & Plunkett, K. (1997). *Rethinking innateness: A connectionist perspective on development (neural networks and connectionist modeling)*. MIT Press.

Gupta, N., Jang, Y., Mednick, S. C., & Huber, D. E. (2012). The road not taken: Creative solutions require avoidance of high-frequency responses. *Psychological Science*, *23*(3), 288 – 294.

Hills, T. (2013). The company that words keep: comparing the statistical structure of child- versus adult-directed language. *Journal of Child Language*, *40*, 586–604.

Jones, M. N., & Mewhort, D. J. K. (2007). Representing word meaning and order information in a composite holographic lexicon. *Psychological Review*, *114*, 1–37.

Kajić, I., & Wennekers, T. (2015). Neural network model of semantic processing in the remote associates test. In *Workshop on Cognitive Computation: Integrating Neural and Symbolic Approaches, 29th Annual Conference on Neural Information Processing Systems (NIPS 2015).*

Klein, A., & Badia, T. (2015). The usual and the unusual: Solving remote associates test tasks using simple statistical natural language processing based on language use. *The Journal of Creative Behavior*, *49*(1), 13–37.

Kounious, J., & Beeman, M. (2014). The cognitive neuroscience of insight. *Annual Review of Psychology*, *65*, 71–93.

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*(2), 211.

Mednick, S. A. (1962). The associative basis of the creative process. *Psychological Review*, *69*(3), 220-232.

Michel, J.-B., Shen, Y. K., Aiden, A. P., Veres, A., Gray, M. K., Pickett, J. P., ... others (2011). Quantitative analysis of culture using millions of digitized books. *Science*, *331*(6014), 176–182.

Nelson, D. L., McEvoy, C. L., & Schreiber, T. A. (2004). The University of South Florida free association, rhyme, and word fragment norms. *Behavior Research Methods, Instruments, & Computers*, *36*(3), 402-407.

Powell, A., & Vega, M. (1971). Word association and verbal analogy problems. *Psychonomic Science*, *22*(2), 103–104.

Stewart, T. C., Bekolay, T., & Eliasmith, C. (2011). Neural representations of compositional structures: Representing and manipulating vector spaces with spiking neurons. *Connection Science*, *22*, 145-153.

Steyvers, M., Shiffrin, R. M., & Nelson, D. L. (2004). Word association spaces for predicting semantic similarity effects in episodic memory. *Experimental Cognitive Psychology and its Applications: Festschrift in Honor of Lyle Bourne, Walter Kintsch, and Thomas Landauer*, 237–249.

Toivonen, H., Gross, O., Toivanen, J. M., & Valitutti, A. (2013). On Creative Uses of Word Associations. In *Synergies of Soft Computing and Statistics for Intelligent Data Analysis* (Vol. 190, p. 17-24). Springer Berlin Heidelberg.