Inferring Metaphoric Structure from Financial Articles Using Bayesian Sparse Models

Martin Sälzle (saelzle_martin@ceu-budapest.edu)
Department of Cognitive Science, Central European University
Frankel Leó útca 30-34, Budapest, 1023, Hungary

Mark T. Keane (mark.keane@ucd.ie)
School of Computer Science & Informatics, University College Dublin
Belfield, Dublin 4, Ireland

Abstract

Drawing from a large corpus (17,000+ articles) of financial news, we perform a Bayesian sparse model analysis of the argument-distributions of the UP and DOWN-verbs, used to describe movements in indices, stocks and shares. Previous work, by Gerow and Keane (2011a, 2011b, 2011c), has shown, using measures of overlap and k-means clustering, that metaphor hierarchies and antonymic relations can be found in this data; for instance, UP verbs have *rise* as a superordinate organizing a distinct set of subordinate verbs (*soar*, *jump*, *climb*, *surge*, *rebound*, *advance*). This work empirically realizes theories about the structuring of our conceptual systems with metaphors (Lakoff, 1992; Lakoff & Johnson, 1980) but does so using a distributional approach to meaning; namely, that words that occur in similar contexts have similar meanings (see Wittgenstein, 1953). However, Gerow and Keane's analysis does not show the overall structure of how these metaphors semantically relate to one another. In the present paper, we re-analyzed their data using a Bayesian sparse model (Lake & Tenenbaum, 2010) in order to infer this metaphor space as a uniform representation, based on the argument distributions. Therefore, we treated arguments as features of metaphors. Our model learned three dimensional graphs in an unsupervised manner as sparse representations of the metaphoric structure over all argument distributions, in parallel. Doing so, it also successfully indicates the metaphoric hierarchies and antonymy relations, that were found by the previous models. In conclusion, we discuss the benefits of this approach.

Keywords: Argument features; analogy; Bayesian inference; emergent structure; corpus analysis; metaphor hierarchies; semantic cognition; similarity; sparse representation; spatial metaphor; structure discovery; unsupervised learning.

Introduction

In recent years, significant progress has been made in deriving meaning from statistical analyses of distributions of words (e.g., Gerow & Keane, 2011a; Landauer & Dumais, 1997; Turney, 2006; Turney & Pantel, 2010; Michel et al., 2010). This distributional approach to meaning takes the view that words that occur in similar contexts tend to have similar meanings (see Wittgenstein, 1953) and that by analyzing word usage we can recover meaning. For instance, Michel et al., (2010) argue that significant insights into human culture and behavior can be derived from analyzing very large corpora, such as the GoogleBooks repository.

Gerow and Keane (2011a-c; henceforth abbreviated as G&K) took such a distributional approach to understanding metaphorically-structured knowledge (in hierarchies and antonymic relationships) between “UP" and "DOWN” verbs from a corpus of financial news reports. Lakoff and Johnson (1980) have argued that metaphors are used to structure many domains of human experience and also many abstract conceptual domains (e.g., emotions). They specifically identified the use of the UP-DOWN metaphor opposition in accounts of wealth (e.g., WEALTH-IS-UP as in *high class*) and in the *rise* and *fall* of numbers (e.g., MORE-IS-UP; LESS-IS-DOWN).

G&K (2011a) build a corpus of 17,000+ financial articles covering a 4-year period, about the major world stock indices (Dow Jones, NIKKEI, FTSE-100) from the *Financial Times*, *NY Times* and *BBC* websites; the corpus contained over 10M words. After parsing the corpus, G&K selected all the sentential instances of the most commonly occurring UP and DOWN verbs (see G&K, 2011a, 2011b for details). Table 1 shows some of the most commonly used arguments found in the corpus, indicating the metaphoric usage of the selected verbs. G&K then analyzed the clustering in these distributions (using k-means clustering) and the overlaps between the distributions of different verbs (using the % overlap in each pair-wise comparison of verb arguments). This analysis threw up some striking regularities.

Table 1: The percentage of *rise's* argument distribution covered each of the 10 most frequent arguments.

Rank	Argument Word	% of Corpus
1	Index	7.3
2	Share	5.6
3	Point	4.8
4	Percent	2.9
5	Price	2.4
6	Stock	2.0
7	Yield	1.9
8	Cent	1.3
9	Profit	0.9
10	Rate	0.9

Metaphor Hierarchies

G&K (2011b) argued that if one verb-metaphor (e.g., that referred to by *rise*) was organizing another metaphoric verb (e.g., *soar*) then the argument distribution of the former should largely cover the latter, but the opposite would not be the case. They also argued that verb-metaphors at the same level of generality (e.g., a basic level), *sibling metaphors*, would have symmetrically overlapping argument distributions. Their coverage- and cluster analysis confirmed these types of structure. On coverage, they found that *rise* organized a group of metaphoric siblings (*soar, jump, climb, surge, rebound, advance*), set off from a set of other more outlying verb-metaphors (*increase, rally, recover, gain, alleviate, elevate*; see Figure 1, A). In the clustering, they found that *rise* was quite separate from all the other verbs that clustered together and that *gain* and *climb* were quite distinct (see Table 2). A similar pattern was found for *fall* and its subordinate-related verb-metaphors (*dip, retreat, sink, plunge, tumble, slide slip, slide, ease, drop, ease, plummet*), with *decline, lose, decrease* and *worsen* being outliers (see Figure 1, B, and G&K, 2011b, for detailed results).

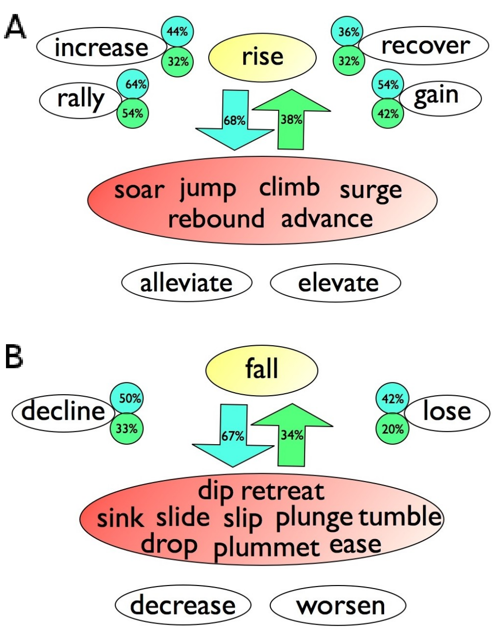

Figure 1: Argument coverage of (A) main UP-verbs and (B) main DOWN-verbs from G&K (2011b).

Metaphoric Antonyms

G&K (2011c) also analyzed verb distributions for antonymic relations; arguing that preferred antonyms *rise-fall* should have more similar distributions than less preferred antonyms *rise-decrease* or *rise-lower*. G&K performed a psychological experiment to find the preferred antonyms between the UP and DOWN verbs and then formulated several different similarity measures (Euclidean distance, cosine similarity, K-L divergence) on the argument distributions to determine which one best predicted the human choices. Given a set of 13 UP verbs and 15 DOWN verbs (as possible antonyms) people identified 114 unique antonym pairs. Of these, in 60% of cases, the cosine similarity of the argument distributions of pairs correctly identified the most preferred antonym-pair from the human ratings. This figure rose to 87% if we consider identifying the 1^{st} or 2^{nd} most preferred pairs (see G&K, 2011c for details). Table 3 lists some results of the human antonymy ratings.

Table 2: Top 5 clusters in k-means analysis of UP-verbs (* rest = the remaining verbs).

Rank	Cluster Groups	% of Tot. (Freq.)
1	rise, rest*	62% (1451)
2	rise, gain, rest*	18% (702)
3	rise, [climb, gain], rest*	4% (36)
4	rise, [jump, climb, gain], rest*	3% (27)
5	all-verbs-as-one-group	2% (18)

Table 3: Some examples of people's verb antonymy ratings, conducted by G&K (2011c). Percentage measures indicate mean antonymy ratings over participants and sub-tasks (free generation and match the opposite).

Verb pair	Antonymy
rise-fall	57%
jump-fall	31%
drop-climb	13%
decline-rise	27%
slide-climb	23%
soar-plummet	17%

Using Sparse Models Instead

G&K found a number of interesting regularities for hierarchical and antonymic relationships between the argument distributions of UP and DOWN verbs. However, their results were based on different approaches, rather than a unifying model, and do not indicate the semantic structure of the metaphoric corpus as a whole. Arguably, it is essential to understand the cognitive semantics of the corpus, as the meaning of individual concepts must depend on how they relate to one another (Kemp & Tenenbaum, 2008). Bayesian sparse models, also known as sparse graph codes (MacKay, 2003), appear to be good candidates for this task.

Bayesian sparse models basically infer an emergent structure in a probabilistic framework (Rogers & McClelland, 2004). Applied to semantics, they have been shown to perform particularly well at finding regularities for the clustering of features for very large numbers of words from different conceptual domains (Lake & Tenenbaum, 2010). These models assume that people learn a set of parameters that fit their observed data well.

Sparse models may be better at handling metaphoric structure than other structured probabilistic models for semantic cognition (e.g., Kemp & Tenenbaum, 2008). The latter generate structures as instances of forms and discover the structural instance of the form that best explains the underlying dataset; including, structural instances based on the graph grammar of trees, linear orders, multidimensional spaces, rings, dominance hierarchies, cliques, and other forms that are supposed to be the organizing principles for data of different cognitive domains. In this way, these models account for domain-specific inferences. The learned structures can then be used to model human inductive reasoning about novel properties of objects within those domains (Kemp & Tenenbaum, 2009).

However, considering a dataset of metaphors, we need to take into account that contemporary cognitive linguistics understands conceptual metaphor not as domain-specific inference but rather as mappings from one conceptual domain to another (Lakoff, 1987; Gibbs, 1994, 1996; Fauconnier & Turner, 1998, 2003). For example, mapping the directionality of movement to changes in quantity (e.g., "prices are rising"). A cognitive model of metaphoric structure would, therefore, not necessarily need to select between structural instances of domain specific forms. Since a metaphoric corpus is likely to consists of many mappings of many different conceptual domains, it would rather need to infer an emergent structure on the basis of a psychologically justified prior probability over the hypothesis space of possible structures. We think that sparseness would be a useful prior for such a model, as it accounts for the cognitive parsimony that is needed to mentally structure metaphors over the vast array of semantic domains (Lakoff, 1992); as well as for the trade off between cognitive effect and computational effort (Wilson & Carston, 2006).

Sparse Model Analysis of Verb-Metaphors

Lake and Tenenbaum's (2010) Bayesian sparse model was used on G&K's verb-metaphor corpus, involving UP and DOWN verbs, extracted from the larger finance corpus (see Data Set). This metaphor corpus contained 9,700+ distinct sentence instances for these two sets of verbs. The sparse model should be able to learn and graph the structure of these verb-metaphors by determining how they covary with regard to the frequency of their argument features. Graphically, the verb-metaphors are represented as nodes in a weighted graph, where the strength of the link between two object-nodes is related to the weighted covariation of their features. The weights of the graph, denoted as the symmetric matrix W, are learned from data by optimizing an objective function that trades off the fit to the data with the sparsity of the graph. In the present study, the sparse model technique was used to build three different graphs: a graph for the UP verb-metaphors and their arguments (using a 13 x 386 matrix), a graph for the DOWN verb-metaphors and their arguments (using a 15 x 456 matrix), and a graph of the combined set of UP and DOWN verb-metaphors (using a 28×605 matrix).[1]

Method

Data Set A total of 28 verbs were used, 13 UP-verbs with 386 distinct, unique arguments, 15-DOWN verbs with 456 distinct, unique arguments (based on those used by G&K, 2011b-c). There were 9,721 distinct sentence instances in the corpus (5803 sentences with UP verbs, 3918 sentences with DOWN verbs).

Model Setup The code for the model we used was written in MATLAB (provided by Brenden Lake, Department of Brain and Cognitive Sciences, MIT; see Lake and Tenenbaum, 2010, for a detailed description of the model). Formally, the undirected graph W defines a multivariate Gaussian distribution $p(f^{(k)}|W)$ in the generative model, known as a Gaussian Markov Random Field (GMRF), where the n objects are the n-dimensions of the Gaussian. With a prior distribution on sparsity, the model then estimates the maximum a posteriori (MAP) parameters W as optimal structure based on data. Each data set D was cast in a $n \times m$ matrix with n metaphors and m arguments. Therefore, the columns of D, denoted as arguments $\{f^{(1)}, ..., f^{(m)}\}$, were assumed to be independent and identically distributed drawn from $p(f^{(k)}|W)$. With the n-dimensional Gaussian distribution, it is assumed that arguments vary smoothly over the graph. So, if two metaphors i and j happen to be connected by a large weight (w_{ij}), they share similar application frequencies over arguments. As a result of sparsity, most metaphors are not directly connected in the learned graph (i.e., w_{ij}=0). The resulting weights allowed us to further apply a Markov Cluster Algorithm (MCA) to classify verb metaphors based on the covariation of their argument distributions. Inflation and pre-inflation settings for the MCA were hold on standard (see Freeman et al., 2007; Theocharidis, Van Dongen, Enright, & Freeman, 2009).

Results & Discussion

Figure 2 shows the resulting weight matrices illustrated as sparse graphs learned for the three different datasets: (A) UP verbs, (B) DOWN verbs and (C) UP-DOWN verbs combined (all graphs were drawn with BioLayout Express 3D; videos of rotating versions of respective graphs should be retrievable by clicking on them). In each graph, the labeled nodes represent verb-metaphors (e.g., *rise, fall*). The links show the connection weights and consequential distances between the nodes, denoting similarity over all

[1] Resulting weight matrices are available from the authors.

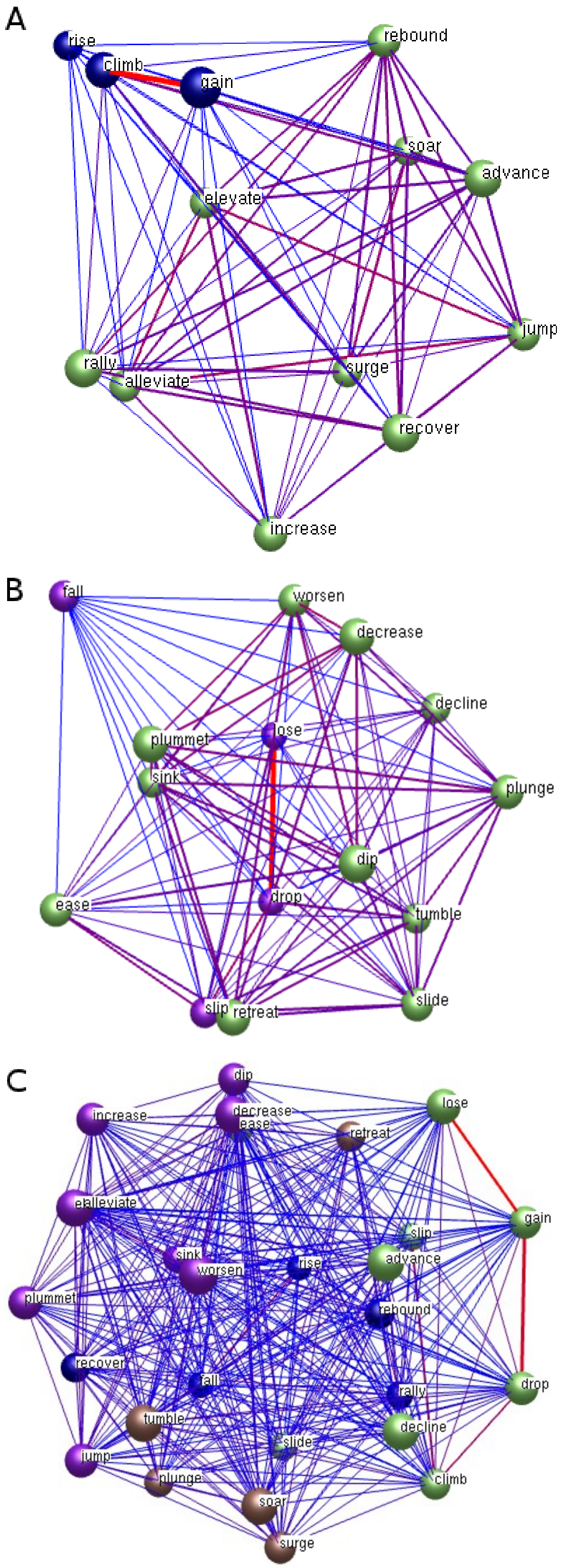

Figure 2: Resulting sparse graphs for (A) UP-metaphors, (B) DOWN-metaphors, and (C) UP-DOWN metaphors.

verbs graphed. The color and thickness of the links represent the weight magnitude; red links indicate high weights (with thickness indicating higher weights), whereas blue links indicate low weights (between 0 and 1). Colors of nodes denote classes of verb-metaphors found by the MCA.

Metaphoric Structure of UP Verbs Figure 2 (A) shows the sparse graph for the UP verb-metaphors. Overall, it literally provides a much better picture of the semantic space of the metaphors with the relative distances between each clearly shown, compared to G&K's (2011b, 2011c) analyses. First, note that the *rise* node stands out as being distinct and non-similar to most of the other nodes. Counter-intuitively, this occurs because though *rise* has arguments that cover many of the arguments of most other verbs *combined* (also see Figure 1, A), it has fewer arguments in common *individually* with any given verb (and, therefore, low similarity with each). Second, *rise*, *climb* and *gain* cluster separate to the remaining verb-metaphors (purple vs. green nodes). While we know that *rise* has asymmetric coverage regarding most other verbs, *climb* and *gain* have not (also see Figure 1, A). Therefore, the latter are two highly interconnected outliers.

Metaphoric Structure of DOWN Verbs Figure 2 (B) shows the sparse graph for the DOWN verb-metaphors. Here, again, the results are very similar to what we saw for the UP verbs. Again, the sparse model analysis provides a much better picture of the semantic space of the metaphors with the relative distances between each clearly shown. First, note that now the *fall* node stands out as being distinct and non-similar to most of the other nodes; for the same reasons we advanced for *rise*. However, regarding the coverage, it can just be considered as a superordinate to some of the other verbs (see Figure 1, B). Second, there is a marked cluster of verbs that are all (almost) equally similar to one another (green nodes). Third, there is another set of verbs that are similar but distinct (*lose*, *drop* and *slip*). While *slip* is an outlier, *lose* and *drop* are highly similar. So, again, while these graphs give a better picture of the space, they may need to be supplemented by coverage measures in defining whether nodes might be actual superordinates or simply unrelated.

Metaphoric Verb Antonyms Figure 2 (C) shows the sparse graphs for the combined UP and DOWN verb corpora. These graphs are slightly different because they deal with both categories of verbs. G&K's (2011c) analysis for antonymy worked on the basis that the key antonyms would be highly similar, relative to other pairings across the two sets of verbs. Again, the sparse model shows this very clearly as notable key antonym pairs appear as close nodes: for instance, *rise-fall, gain-drop, climb-slip, gain-lose.* Further, how the verb-metaphors cluster (shown by node coloration) indicates semantic similarity in how they got applied. However, the antonymy ratings from the human subject experiment of G&K (2011c) correlate just weakly with the ones from the model (Pearson's r=0.4; see Figure 3). This might have experimental- and model related reasons: first, the verb-metaphors from the corpus were applied by human speakers to describe financial changes. The experimental data, however, are abstract antonymy ratings of verbs, having neither applicational relation to the domains relevant to conceptualize finance, nor to any other cognitive domain. (Future experiments for metaphoric antonymy would need to take this into account.) Second, the model's antonymy ratings for the superordinates *rise* and *fall* had to be excluded, since they were 0 to all other metaphors, except to one another. Finally, in the graph are also some high-similar pairings within the same verb set, like *rally-rebound* and *slip-ease*, that are clearly "just similar" and not antonyms. The latter indicates that some prior categorization of what-are-known-to-be broadly opposite sets is required before such a merged model might be useful. Again, an additional coverage analysis is needed to isolate *rise* and *fall* as superordinates (see also Figure 1, A and B).

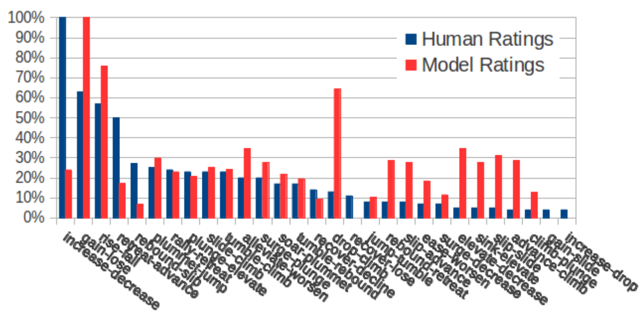

Figure 3: Model- versus human antonymy rating of verb-pairs in per cent. Human ratings reflect perceived antonymy of verbs (see G&K, 2011c); whereas model ratings reflect the computed antonymy of verb-metaphors used to describe financial changes. The latter are weighted entries of the model's weight matrix.

Conclusion

The suggestion that significant parts of our conceptual systems are structured by metaphors has mainly received support from linguistic and anthropological analyses (see Lakoff & Johnson, 1980). However, cashing out these ideas empirically in a systematic way has proven difficult. The promise of the present work is that these ideas can be empirically supported by a distributional analysis of verb arguments, with such metaphoric import. We have shown that sparse models can provide a rich and informative basis for relating these verb-metaphors together in a uniform metaphor space. We believe that this approach may be useful in modeling other cognitive tasks that rely on these metaphoric spaces (e.g., language comprehension, analogical thinking). For instance, in analogical thinking it has long been argued that conceptual slippage (Hofstadter, 1995) and re-description (Keane, 1996; Kurtz, 2006) are needed to account for human abilities: Bayesian sparse models provide a basis for allowing such slippage, assuming structural support for the slippage being considered.

However, our work has also indicated that the sparse models will still need a coverage analysis to isolate superordinate metaphors. And, because these are important for conceptually structuring the metaphor space, they should be implemented in the way sparse models generate and learn structure. This might be achievable by using hierarchical Bayesian sparse models (Chandrasekaran, Parrilo, & Willsky, 2010) that potentially discover organizing metaphoric concepts as hidden or latent variables, and further increase sparsity.

Acknowledgments

This work was carried out as part of a self-funded MSc in Cognitive Science at UCD by the first author. We want to thank Brenden Lake for providing the code for the sparse model, Aaron Gerow for all the data, Anne Tamm for linguistic advice, and Máté Lengyel for technical suggestions.

References

Chandrasekaran, V., Parrilo, P. A., & Willsky, A. S. (2010). Latent variable graphical model selection via convex optimization. In *The 48th Annual Allerton Conference on Communication, Control, and Computing*, (pp. 1610–1613). IEEE.

Fauconnier, G., & Turner, M. (1998). Conceptual integration networks. Cognitive science, *22*(2), 133–187.

Fauconnier, G., & Turner, M. (2003). *The way we think: Conceptual blending and the mind's hidden complexities*. Basic Books.

Freeman, T. C., Goldovsky, L., Brosch, M., Van Dongen, S., Mazière, P.,...Enright, A. J. (2007). Construction, visualisation, and clustering of transcription networks from microarray expression data. *PLoS Computational Biology*, *3*(10).

Gerow, A., & Keane, M. T. (2011a). Mining the Web for the"Voice of the Herd" to track stock market bubbles. In Proceedings of the *22nd International Joint Conference on Artificial Intelligence*: Barcelona, Spain.

Gerow, A., & Keane, M. T. (2011b). Identifying metaphor hierarchies in a corpus analysis of finance articles. In

Proceedings of the 33rd Annual Meeting of the Cognitive Science Society: Boston, MA.

Gerow, A., & Keane, M.T. (2011c). Identifying metaphoric antonyms in a corpus analysis of finance articles. In *Proceedings of the 33rd Annual Meeting of the Cognitive Science Society*: Boston, MA.

Gibbs, R. W. (1994). *The poetics of mind: Figurative thought, language, and understanding*. Cambridge University Press.

Gibbs, R. W. (1996). Why many concepts are metaphorical. *Cognition, 61*, 309–319.

Hofstadter, D. R. (1995). *Fluid concepts and creative analogies: Computer models of the fundamental mechanisms of thought.* Together with the Fluid Analogies Research Group. NY: Basic Books.

Keane, M. T. (1996). On adaptation in analogy. *Quarterly Journal of Experimental Psychology, 49A*, 1062-1085.

Kemp, C., & Tenenbaum, J. B. (2008). The discovery of structural form. *Proceedings of the National Academy of Sciences, 105*(31), 10687-10692.

Kemp, C., & Tenenbaum, J. B. (2009). Structured statistical models of inductive reasoning. *Psychological Review, 116*(1), 20-58.

Kurtz, K. J. (2005). Re-representation in comparison: Building an empirical case. *Journal of Experimental & Theoretical Artificial Intelligence, 17*(4), 447–459.

Lake, B., & Tenenbaum, J. (2010). Discovering structure by learning sparse graph. In *Proceedings of the 33rd Annual Cognitive Science Conference*: Boston, MA.

Lakoff, G. (1987). *Women, fire, and dangerous things. What/How Categories Reveal About the Mind.* Chicago: Chicago UP.

Lakoff, G. (1992). *The contemporary theory of metaphor*. In A. Ortony (Ed.), Metaphor and Thought 2nd Edition. Cambridge University Press.

Lakoff, G., & Johnson, M. (1980). *Metaphors We Live By*. Chicago, IL: University of Chicago Press.

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.

MacKay, D. J. C. (2003). *Information theory, inference, and learning algorithms*. Cambridge, UK: Cambridge University Press.

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

Rogers, T. T., & McClelland, J. L. (2004). *Semantic cognition: A parallel distributed processing approach*. Cambridge, MA: MIT Press.

Theocharidis, A., Van Dongen, S., Enright, A. J., & Freeman, T. C. (2009). Network visualization and analysis of gene expression data using BioLayout Express3D. *Nature protocols, 4*(10), 1535–1550.

Tenenbaum, J. B., Griffiths, T. L., & Kemp, C. (2006). Theory-based Bayesian models of inductive learning and reasoning. *Trends in Cognitive Sciences*, *10*(7), 309-318.

Turney, P. D. (2006). Similarity of semantic relations. *Computational Linguistics, 32*(3), 379-416.

Turney, P. D., & Pantel, P. (2010). From frequency to meaning: Vector space models of semantics. *Journal of Artificial Intelligence Research, 37*(1), 141-188.

Wilson, D., & Carston, R. (2006). Metaphor, relevance and the 'emergent property' issue. *Mind & Language*, *21*(3), 404–433.

Wittgenstein, L., & Anscombe, G. E. M. (1953/2001). *Philosophical investigations*: the German text, with a revised English translation. Wiley-Blackwell.