

Words, Concepts, and the Geometry of Analogy

Stephen McGregor

Matthew Purver

Geraint Wiggins

Queen Mary University of London

School of Electronic Engineering and Computer Science

s.e.mcgregor@qmul.ac.uk

m.purver@qmul.ac.uk

geraint.wiggins@qmul.ac.uk

This paper presents a geometric approach to the problem of modelling the relationship between words and concepts, focusing in particular on analogical phenomena in language and cognition. Grounded in recent theories regarding geometric conceptual spaces, we begin with an analysis of existing static distributional semantic models and move on to an exploration of a dynamic approach to using high dimensional spaces of word meaning to project subspaces where analogies can potentially be solved in an online, contextualised way. The crucial element of this analysis is the positioning of statistics in a geometric environment replete with opportunities for interpretation.

1 Concepts in the Kingdom of Context

“Laugh to scorn the power of man, for none of woman born shall harm Macbeth”—so says a phantasmal apparition to the ill-fated anti-hero of Shakespeare’s gruesome tragedy. And, taking the ghost at its word, Macbeth goes merrily on his bloodthirsty, tyrannical way until one day a man born by caesarean section chops him down. What’s at play here is, of course, the context through which the utterance has been interpreted: Macbeth made the classic mistake of taking a demonic entity literally rather than spending a little time thinking about malicious ways in which words describing a seemingly impossible scenario might actually be saying something imminently plausible.

In as much as theoretical linguistics is concerned, context, much more than just the stuff of literary misapprehensions, has become more or less the central problem in terms of understanding the way that words relate to situations in the world. Whether it be type theory [22, 9], event semantics [10], relevance theory [27], or any of a variety of other theoretical approaches to semantics, the ability to map word meaning to a specific mental experience hinges on the situatedness of both cognition and language. What is at stake here is the fraught relationship between words and concepts. Barsalou asks, “What kinds of representations are sufficiently expressive to account for the hierarchical conceptual relationships of concepts?”, [4], and concludes that, just as the labelling of conceptual content is *haphazard*, so language must be characterised by a fundamental *vagary* and *flexibility*.

Yet there is something in the contextualisation of meaning that confounds computational approaches to language. Symbol manipulating machines, bound by the literalness of the formal languages in which they traffic, balk at handling the immensity of context that can be associated with any given symbol in the course of natural communication. An attempt to build robust symbolic representations linking linguistic elements to all their potential contextualised applications more or less immediately encounters a *frame problem* by which a preponderance of potential contexts explodes with even very slight shifts in the situation of a cognitive-linguistic agent in the real world [11]. The question, then, is how computers might use low-level statistical type representations to capture the environmentally situated conceptual dynamism that is essential to cognition.

Gärdenfors has proposed a theory of *conceptual spaces* which moves conceptual modelling into a spatial domain that should at least in principle be more tractable to symbol manipulating machines

[14]. By describing concepts as being situated in spaces where their dimensions can be interpreted as properties of the concept, this theory offers hope that concepts might be modelled in terms of the kind of matrix of quantifications that are native to computers. Widdows similarly proposes that concepts can be understood as geometric phenomena, focusing more explicitly on the relationship between concepts and words and using statistical approaches to lexical modelling to build a framework for logical operations on conceptual representations [29]. In both cases, these approaches to concept modelling offer a kind of interstitial layer of information processing that sits between the low level data of environmental stimuli or textual statistics and the high level symbolism of cognitive perception.

This existing work provides a theoretical grounding for the idea that will be explored in the following pages, where the groundwork for developing a methodology for moving from word counts to conceptual models will be laid. This paper will address a particular type of conceptual relationship: analogy, a conceptual mechanism with a pedigree dating back at least to Aristotle [2]. The utility of analogy as a mechanism for performing conceptualisation in scientific domains has been thoroughly attested [17], and more generally analogy is inherent in the relationships that can be mapped across the hierarchical structure of a taxonomy. So a move from the statistically proximate to the geometrically analogical is a move from the lexical to the conceptual. The crucial aspect of the approach which will be proposed here is its geometry: by maintaining spaces that, in the spirit of Gärdenfors and Widdows, are strictly interpretable right down to the level of a single dimension, we hope to be able to eventually establish a robust computational process for dynamically extracting context sensitive conceptual mappings from large scale textual corpora.

Our dynamic model for projecting context specific, conceptually productive subspaces should likewise be situated in terms of other ongoing technical work in distributional semantics and further afield. In particular, a similar approach has been explored for the computational discovery of taxonomic relationships such as hypernymy and hyponymy [24], and likewise for the generation of spaces where directions in space can be understood semantically [12]. A key feature of each of these approaches is the establishment of a distributional semantic space where the preservation of statistical information about an underlying corpus yields an element of interpretability to the space. More generally, work on knowledge base construction and extension has likewise treated geometric properties of spaces of entities as translations which can be passed between pairs of entities and likewise be interpreted as meaningful features of the space [6].

These approaches involving the translation of statistical information about a corpus into spaces where geometry coincides with conceptual properties should be contrasted with theories that frame concepts in terms of rule based manipulations of content-bearing symbols [13]. The case to be made here is, broadly, that cognition is an inherently representational process, anchored in the intentionality of the symbols in which it traffics rather than in the geometry of conceptual spaces. Ultimately, though, there is hope that the geometrical approach can actually stand in as a construct mitigating between low level stimuli and high level symbolic representation. There is likewise a point of comparison with work where conceptual relationships are modelled as sets of nodes in a semantic network [16]. The latter type of model is characterised by abstract graph theoretical representations where relationships, signified by labelled edges, are inherently symbolic and geometry is not a factor in the interpretation of the semantic situation. Having said this, it must also be noted that there is considerable scope for reconciling graph based approaches to conceptual modelling to the geometric methodology that will be endorsed here [15].

2 Spaces of Meaning

A prevalent paradigm in corpus-based computational approaches to lexical modelling is *distributional semantics*, which seeks to establish spaces where words are treated as vectors and the relative proximity of vectors corresponds to the semantic relatedness of the words that label those vectors. In particular, the *distributional hypothesis* holds that “words that occur in similar contexts tend to have similar meanings,” [28, p. 148], and so the intuition behind distributional semantics is that if the typical textual context of word types can somehow be statistically captured, then there is hope for constructing computationally tractable representations of word meaning based on word-vectors populated by these statistics. Traditional approaches have utilised the frequencies or rates of word co-occurrences to build distributional semantic spaces [25, 26], while more recent work has often employed non-linear regression to learn an optimally arranged space [5, 8] or used matrix factorisation to optimise the informativeness of condensed spaces [3], though, as will be seen in the following analysis, the geometrical character of these spaces can vary. Clark provides a current and comprehensive overview of the field [7].

Of particular relevance to the ideas presented in the present paper is the recent development of the word2vec distributional semantic model, which uses a neural network to learn a space of word meanings based on iterative traversals of a large scale corpus [21]. Rather than building up statistical representations of word co-occurrences where the number of features in a space can be on the scale of or even significantly larger than the set of word-vectors constituting the space’s vocabulary of word-vectors, this model gradually learns a space which ultimately resembles the statistically derived spaces. The network approach to distributional semantics uses a set of arbitrary dimensions to gradually pull the space into order, further utilising negative sampling to push word-vectors that do not bear semantic relatedness apart.

The upshot of this approach is a space which yields some interesting semantic properties which at first appear to move towards the modelling of conceptual relationships. In particular, it has been demonstrated that in word2vec spaces, straightforward linear algebraic procedures can be used to map analogical relationships between words. The paradigmatic example offered in the literature is this:

$$\overrightarrow{woman} - \overrightarrow{man} + \overrightarrow{king} \approx \overrightarrow{queen} \quad (1)$$

To put it slightly differently, the vector which projects from the point in space representing *man* to the point in space representing *woman* is roughly equal in both length and direction to the vector going from *king* to *queen*—or, to attempt to interpret this geometrically, these four word-vectors very closely form a parallelogram sitting in the high-dimensional space of word2vec. The robustness of this methodology for analogy completion has been demonstrated on a fairly large dataset consisting of some tens of thousands of analogies that are input as $\frac{3}{4}$ complete, with the model returning the correct fourth component of the analogy more than half the time [20]. It is worth noting that similar results have subsequently been reported using a model built through a combination of neural networks and matrix factorisation [23], and it has been suggested that the strong performance of word2vec might be down to the fine tuning of a set of hyperparameters which could likewise be implemented in a more traditional statistical distributional semantic model [18].

Nonetheless, it is notable, if not outright surprising, that these types of ostensibly geometric relationships should emerge from a highly non-linear analysis of co-occurrence relationships across a large scale corpus. In particular, at this point, the relationships at least seem to veer away from the purely lexical and into the conceptual. By capturing the relationship between word-vectors, the model begins to flesh out the corresponding relationships between correspondingly labelled concepts: so, for instance, we dis-

cover that in the conceptual domain of ROYALTY, there is a dimension of GENDER. By systematically going through the ostensible geometry of the space, we might imagine gradually building up a network of relationships resembling a kind of taxonomy, where the coherence between labels across domains offers the trace of a space of interrelated ideas about things in the world.

A closer inspection of the model's operation, however, reveals some significant caveats. For one thing, the dataset used in the evaluation of the model's analogy completing performance is large, but tends to feature analogies with a distinct character: of the analogies described as "semantic", which is roughly half of them, most involve proper names, and a subset involve familial relationships. Then about half the test set is designed to explore "syntactic" relationships, and these analogies mainly consist of comparisons between morphologically aligned pairs of words. In the case of proper names, the analogical queries are arguably less hindered by the ambiguity that characterises more general usage. And in the case of the more grammatically oriented analogies, the model might be capturing the fact that word stems are associated with a general semantic context, but specific derivational words are also characterised by a particular grammatical context—with, for example, adjectives more likely to occur in the immediate context of nouns, where adverbs are to be found in the context of verbs.

In order to draw out this point a bit, here are six examples of analogies completed by `word2vec`. In each case, the model is offered three input terms, and it finds a fourth term (underlined) based on the linear algebraic operation between vectors described above. For the sake of analysis, three examples have been picked which do seem intuitively to work, and three which don't seem to work:

1. *Paris* is to *France* as *Finland* is to *Helsinki* ✓
2. *girl* is to *boy* as *prince* is to *princess* ✓
3. *fast* is to *faster* as *big* is to *bigger* ✓
4. *elephant* is to *mouse* as *big* is to *huge* ✗
5. *bank* is to *river* as *shoulder* is to *elbow* ✗
6. *picture* is to *paint* as *story* is to *mexican_viola_viola* ✗

In (4), the model seems to stumble on the conceptually crucial notion of antonymy: in a lexical model based on word co-occurrences, words with opposite meanings are actually very likely to occur in similar contexts, so there is not a directed way to know about the difference between scaling symmetrically between LARGENESS and SMALLNESS versus scaling up within the domain of LARGENESS. Likewise in (5), the model evidently stumbles on a classic example of lexical ambiguity, with both *bank* and *shoulder* being used in a contextually specific way that's perfectly clear to an attuned speaker of the language but that proves inexorably muddled for a model that is only aware of morphological statistics. And in the case of (6), where a plausible completion of the relationship would have been something along the lines of *words* (or perhaps, more idiomatically, *tell*), the model has completely failed to capture the metaphoric aspect of this particular analogy, to moderately humorous effect.

In each of the cases offered, the model is seen to fall short when it comes to the essential contextuality of the analogy being mapped. This belies the fundamentally non-geometric characteristic of this type of space: distances corresponding to word similarity can, on their own, occasionally be useful heuristics for predicting more nuanced conceptual relationships denoted by words, but in general this space of word statistics can't really be mapped to a more productive conceptual space where dimensions take on meaningful properties of their own. In order to move from spaces of words to spaces of concepts, we need to discover a more dynamic way of manipulating the space itself, mirroring the fundamental situatedness of the way that a cognitive agent conceptualises about the world.

3 Putting Word Context in Context

So it is evident, and perhaps, from a theoretical perspective, already somewhat obvious, that context is essential in the mapping of the relationship between words and concepts. The problem with the lexical spaces described thus far, however, is that, while they are very effective at capturing a kind of encyclopedic knowledge about the relationships between words and concepts, they are also essentially *static*. In the case of word2vec in particular, the neural network method used to generate the space is specifically designed to maximise the utility of each of the spaces dimensions, treating the dimensions themselves as nothing more than handles to gradually shift the space into order based on iterative observations of words in context across a corpus. And in the case of other state-of-the-art systems, the use of matrix factorisation techniques such as SVD results in the optimisation of the informativeness of each dimension in the space, but at the expense of the degradation of the informativeness of the dimensions themselves. Thus these spaces are *static* in the sense that there is no real hope of finding a mechanism for further interpreting and manipulating their dimensions in a context sensitive way.

Recent work has suggested an alternative approach to distributional semantics, by which a very high dimensional space of word co-occurrence statistics is learned and left unrefined. This base space can then take sets of words as input and analyse the corresponding word-vectors in order to determine the statistically literal co-occurrence dimensions which best capture the conceptual context of the relationship between the input words. The sparsity of the resulting space becomes its strength, affording the projection of a subspace in which terms become clustered in a conceptually coherent way: the base space is *dynamic* in that it responds to input in an online, situational fashion. The intuition behind this approach is illustrated in Figure 1. Furthermore, the approach has been implemented and its conceptual clusterings verified by human subjects [1], in addition to being shown to perform at least equivalently to word2vec on a computational task involving contextualising inputs in the form of compound nouns [19]. It should also be noted that a similar approach has been explored for the computational discovery of taxonomic relationship such as hypernymy and hyponymy [24].

The novel proposal offered in the present paper is that a similarly context sensitive approach involving online projections from a base space should be applicable to the task of generating subspaces where analogical relationships can be mapped in a robustly geometric way. It stands to reason to assume that such a dynamic model should, in principle, be able to find a way, given an analogy, to project some subspace where the analogy can be described as a parallel relationship between paired vectors.

But, whereas the published work mentioned above analyses input by searching for dimensions that maximise the values of a set of input terms, the way forward for projecting analogically productive subspaces must lie in searching for dimensions that optimise the desired geometric relationship between the components of the analogy. To express this formulaically, given an analogy A is to B as C is to D , dimensions should be sought where the values of the analogical components play out such that $A - B \approx C - D$. The hypothesis offered here, as a preliminary step towards an approach to analogy completion and perhaps even metaphor generation, is that dimensions selected on this basis should produce spaces where the previously described geometric relationship between analogical terms emerges.

In order to test this hypothesis and further explore the nature of analogically geared subspaces, the following procedure is proposed, based on an input consisting of an analogy of the form $A : B :: C : D$. First, in line with existing work in contextually dynamic distributional semantics [1, 19], a general base space of co-occurrence statistics is built—in particular, the space used to generate the following data is based on the English language pages of Wikipedia, and the statistical metric employed is pointwise mutual information measuring the likelihood of words co-occurring within 5 words of one another. Specifically, a matrix of word-vectors is calculated as follows, where $n_{w,c}$ is the frequency of observed co-

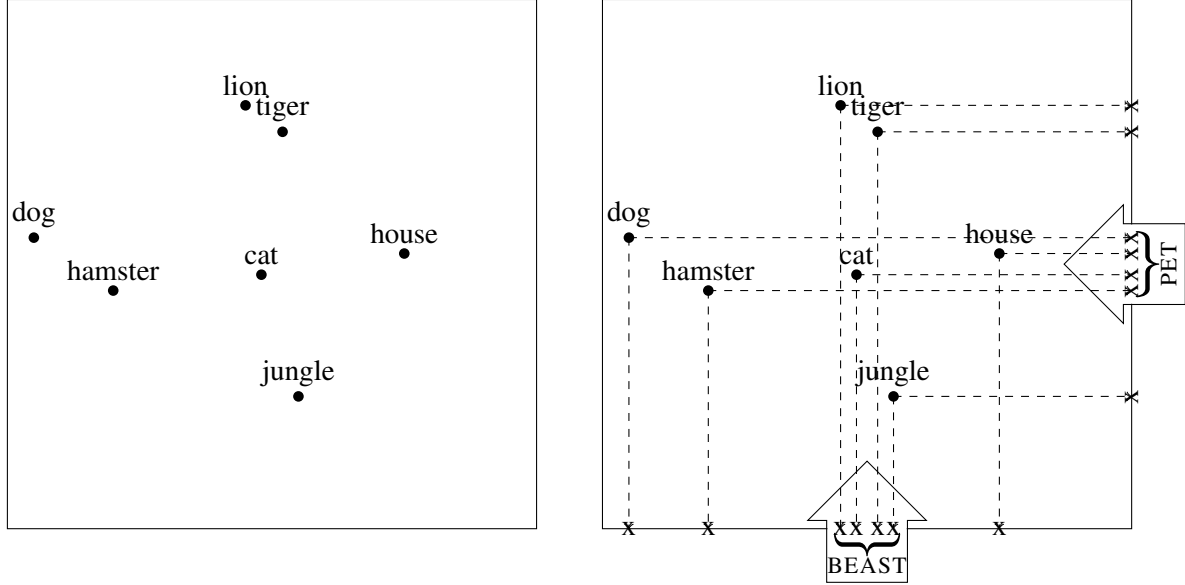


Figure 1: Contextual Projections. On the left, a base space exhibits vagary regarding the concept CAT, captured by the proximity of the word-vector \vec{cat} to terms associated with different connotations of the word *cat*. On the right, two different perspectives on this space reveal context specific conceptual clusterings.

occurrences of context word c within a 5-word window of target word w , n_w and n_c are the independent frequencies of w and c , W is the total number of occurrences of words within the vocabulary, and a is a smoothing constant (set in this case to 10,000, a value determined experimentally):

$$M_{w,c} = \log_2 \left(\frac{n_{w,c} \times W}{n_w \times (n_c + a)} + 1 \right) \quad (2)$$

The result is a sparse space consisting of a vocabulary of word vectors representing the 200,000 most frequent words in Wikipedia arrayed across dimensions representing co-occurrences with approximately 7.5 million word types. With this base space established, the four components of the analogy are analysed, and an initial subspace N of the d dimensions that have non-zero values for the word vectors associated with A, B, and C is built. This dense $3 \times d$ space is then normalised using L1 normalisation, ie, so that the values of each word-vector sum to 1. Next, the 200 dimensions with the highest mean values are selected. In other words, we select the dimensions with the highest value for the following equation:

$$\mu_c = \frac{1}{3} \sum_{w \in \{A,B,C\}} M_{w,c} \quad (3)$$

This step generates a subspace of dimensions that are particularly relevant to the context of the analogy being examined. Furthermore, the normalisation of the $3 \times d$ matrix of non-zero values ensures that a word-vector characterised by dimensions with lower values will still be proportionally represented in the next step of the selection of an analogical context. It's worth noting, as well, that the dimensions selected here are chosen only based on their relevance to the first three components of the analogy, so there is already an element of restriction here: there is no guarantee that component D will be represented

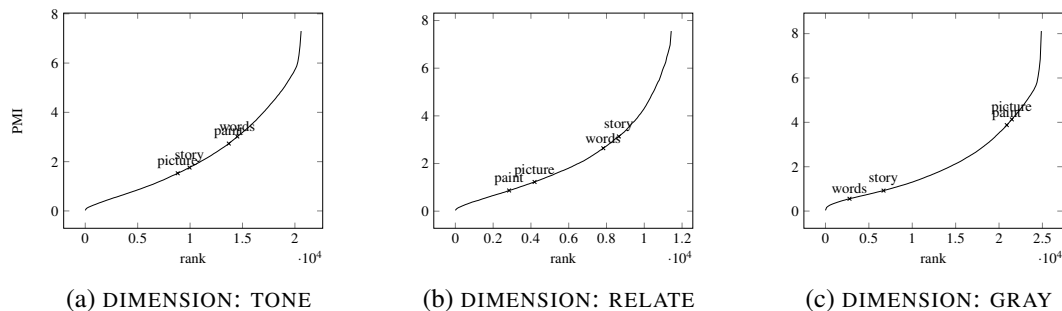


Figure 2: Selected Analogical Dimensions. These are 3 of 20 dimensions that geometrically map the analogy *picture* is to *paint* as *story* is to *words*, presented as histograms ranking word-vectors with non-zero PMI values and highlighting the positions of the analogical components along this curve. Of particular note is the way that analogically oriented word pairs tend to cluster.

with a non-zero value on any of these dimensions. With an eye towards eventually developing a robust methodology for analogy completion, a strong outcome using this procedure guarantees that there is some hope of using only the information afforded by A, B, and C to find an analogical subspace where $A:B::C:D$ is mapped as a predictable geometric relationship.

Next, from these 200 dimensions, the 20 dimensions c with the lowest values for the equation $((M_{A,c} - M_{B,c}) - (M_{C,c} - M_{D,c}))^2$ are selected. This 20-dimensional subspace is then considered the context in which the analogy is mapped.

To demonstrate the robustness of this approach, which is to say, to show that there is almost always *some* subspace where a given analogy can be mapped geometrically, a portion of the word2vec analogy testset consisting of 1,000 analogies has been held back, and a 20 dimensional subspace has been determined for each analogy following the procedure described above. In each case, the space has then been examined to see if the previously attested linear algebraic method for analogy completion was employed, leaving out the fourth term in each case and instead returning the word-vector nearest to the point described by the equation $\vec{B} - \vec{A} + \vec{C}$. Of the 1,000 analogies tested, 3 contained words not in our model's 200,000 word vocabulary; of the remaining 997, only one was not completed correctly.

To further explore the nature of the subspaces where these conceptual relationships are mapped, the previously discussed analogy *picture* is to *paint* as *story* is to *words* has been examined, and a selection of some of the dimensions ultimately used to map the relationship are depicted in Figure 2. The thing to note about the way that the analogy plays out along these dimensions is the way that the relative PMI values of the word-vectors in question cluster into pairs. These pairs will consist of iterations of a given word-vector being coupled with one its two analogues, and over the course of several dimensions a given pair will be observed relatively towards both the high end or the low end of the spectrum of values.

The net geometric result of this tendency is that the parallelogram delimiting the analogy is pushed into an orientation where it sits more or less obliquely near the centre of the positive region of the subspace, approximately perpendicular to a centre line extending from the origin, as illustrated in Figure 3. The distinguishing features of this characteristic analogical geometry – its flatness, its centrality, and its obliqueness – are hardly surprising when considered in the context of the properties of the dimensions that give the edges of the analogy their fundamental trait of parallelism. Nonetheless, the insight gained from pursuing the geometry of these intuitions might be very valuable tools in the geometric description of analogy. This insight might be pushed in at least two different directions. For one thing, given three of the four legs of an analogy, we now have a hunch about the way those three word-vectors should sit

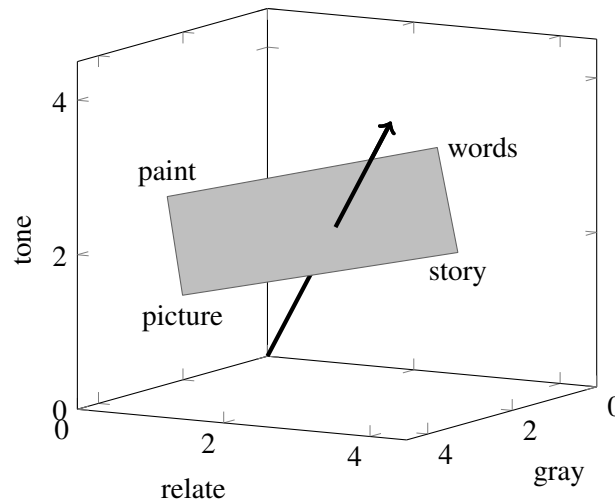


Figure 3: An Analogy in Space. When the analogue values of the three dimensions described in Figure 2 are plotted, the analogy itself emerges as a parallelogram situated obliquely in the centre of the positive region of the space.

in a subspace that will complete the analogy in a satisfactory way. For another thing, given a particular context delineated in terms of co-occurrence dimensions, we now have a mechanism for beginning to look for regions of the space that can bear an analogical conceptual relationship.

4 Final Thoughts

What has been presented here is in effect the beginning of a work-in-progress. The exploration of the way that a dynamic distributional semantic space might be used to project analogically productive subspaces in an online, context sensitive way has not yet yielded a methodology for actually performing, for instance, an analogy completion task. What has been demonstrated, however, is the way that a robustly geometrical approach to corpus linguistics can bear some strong results. Rather than having a space simply for the sake of situating words in relation to one another, we have proposed building spaces which are themselves full of meaning and the potential for interpretation. By situating words in a dynamic space, the problems inherent in the construction of highly informative systems of interactive symbols becomes the natural outcome of spaces which are necessarily replete with interrelationships. So, finally, rather than needing to try to model conceptualisation as a catastrophically unwieldy network of encyclopedic relationships, the cognitive situation inherent in conceptualisation becomes, like the world itself, thoroughly geometric.

References

- [1] Kat Agres, Stephen McGregor, Matthew Purver & Geraint Wiggins (2015): *Conceptualising Creativity: From Distributional Semantics to Conceptual Spaces*. In: *Proceedings of the 6th International Conference on Computational Creativity*, Park City, UT. Available at http://computationalcreativity.net/iccc2015/proceedings/5_4Agres.pdf.
- [2] Aristotle (1895): *The Poetics*. Macmillan and Co, London.

- [3] Marco Baroni & Roberto Zamparelli (2010): *Nouns are vectors, adjectives are matrices: Representing adjective-noun constructions in semantic space*. In: *Proceedings of the 2010 Conference on Empirical Methods in Natural Language Processing*, pp. 1183–1193. Available at <http://dl.acm.org/citation.cfm?id=1870658.1870773>.
- [4] Lawrence W. Barsalou (1993): *Theories of Memory*, chapter Flexibility, Structure, and Linguistic Vagary in Concepts: Manifestations of a Compositional System of Perceptual Symbols. Lawrence Erlbaum Associates, Hove.
- [5] Yoshua Bengio, Rjean Ducharme, Pascal Vincent & Christian Jauvin (2003): *A Neural Probabilistic Language Model*. *Journal of Machine Learning Research* 3, pp. 1137–1155. Available at <http://dl.acm.org/citation.cfm?id=944919.944966>.
- [6] Antoine Bordes, Nicolas Usunier, Alberto Garcia-Duran, Jason Weston & Oksana Yakhnenko (2013): *Translating Embeddings for Modeling Multi-relational Data*. In C. J. C. Burges, L. Bottou, M. Welling, Z. Ghahramani & K. Q. Weinberger, editors: *Advances in Neural Information Processing Systems* 26, Curran Associates, Inc., pp. 2787–2795. Available at <http://papers.nips.cc/paper/5071-translating-embeddings-for-modeling-multi-relational-data.pdf>.
- [7] Stephen Clark (2015): *Vector Space Models of Lexical Meaning*. In Shalom Lappin & Chris Fox, editors: *The Handbook of Contemporary Semantic Theory*, Wiley-Blackwell, doi:10.1002/9781118882139.ch16.
- [8] Ronan Collobert & Jason Weston (2008): *A Unified Architecture for Natural Language Processing: Deep Neural Networks with Multitask Learning*. In: *Proceedings of the 25 th International Conference on Machine Learning*, doi:10.1145/1390156.1390177.
- [9] Robin Cooper (2012): *Type Theory and Semantics in Flux*. In Ruth Kempson, Tim Fernando & Nicholas Asher, editors: *Philosophy of Linguistics*, Elsevier, doi:10.1016/B978-0-444-51747-0.50009-3.
- [10] Donald Davidson (1969): *The Individuation of Events*. In Nicholas Rescher, editor: *Essays in Honor of Carl G. Hempel*, pp. 216–234, doi:10.1007/978-94-017-1466-2.11.
- [11] Daniel C. Dennett (1984): *Cognitive Wheels: The Frame Problem of AI*. In Christopher Hookway, editor: *Minds, Machines, and Evolution: Philosophical Studies*, Cambridge University Press, pp. 129–151.
- [12] Joaquín Derrac & Steven Schockaert (2014): *Characterising Semantic Relatedness Using Interpretable Directions in Conceptual Spaces*. In: *2nd European Conference on Artificial Intelligence*, pp. 243–248, doi:10.3233/978-1-61499-419-0-243.
- [13] Jerry A. Fodor & Zenon W. Pylyshyn (1988): *Connectionism and Cognitive Architecture: A Critical Analysis*. *Cognition* 28(1-2), pp. 3–71, doi:10.1016/0010-0277(88)90031-5.
- [14] Peter Gärdenfors (2000): *Conceptual Space: The Geometry of Thought*. The MIT Press, Cambridge, MA.
- [15] Peter Gärdenfors (2014): *The Geometry of Meaning: Semantics Based on Conceptual Spaces*. The MIT Press, Cambridge, MA.
- [16] Martin Haspelmath (2003): *The geometry of grammatical meaning: semantic maps and cross-linguistic comparison*. In Michael Tomasello, editor: *The new psychology of language: cognitive and functional approaches to language structure*, vol. 2, Lawrence Erlbaum, Mahwah, New Jersey/Londo, pp. 211–242. Available at <http://wwwstaff.eva.mpg.de/~haspelmt/SemMaps.pdf>.
- [17] Mary B. Hesse (1963): *Models and Analogies in Science*. Sheed and Ward, New York.
- [18] Omer Levy & Yoav Goldberg (2014): *Linguistic Regularities in Sparse and Explicit Word Representations*. In: *Eighteenth Conference on Computational Natural Language Learning*, doi:10.3115/v1/W14-1618.
- [19] Stephen McGregor, Kat Agres, Matthew Purver & Geraint Wiggins (2015): *From Distributional Semantics to Conceptual Spaces: A Novel Computational Method for Concept Creation*. *Journal of Artificial General Intelligence*, doi:10.1515/jagi-2015-0004.
- [20] Tomas Mikolov, Kai Chen, Greg Corrado & Jeffrey Dean (2013): *Efficient Estimation of Word Representations in Vector Space*. In: *Proceedings of ICLR Workshop*. Available at <http://arxiv.org/abs/1301.3781>.

- [21] Tomas Mikolov, Wen tau Yih & Geoffrey Zweig (2013): *Linguistic Regularities in Continuous Space Word Representations*. In: *Proceedings of the 2013 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pp. 246–251. Available at <http://www.aclweb.org/anthology/N13-1090>.
- [22] Richard Montague (1974): *English as a Formal Language*. In Richard H. Thompson, editor: *Formal Philosophy: selected papers of Richard Montague*, Yale University Press, New Haven, CT.
- [23] Jeffrey Pennington, Richard Socher & Christopher D. Manning (2014): *GloVe: Global Vectors for Word Representation*. In: *Conference on Empirical Methods in Natural Language Processing*, doi:10.3115/v1/D14-1162.
- [24] Laura Rimell (2014): *Distributional Lexical Entailment by Topic Coherence*. In: *Proceedings of the 14th Conference of the European Chapter of the Association for Computational Linguistics*, Gothenburg, doi:10.3115/v1/E14-1054.
- [25] G. Salton, A. Wong & C. S. Yang (1975): *A Vector Space Model for Automatic Indexing*. In: *Proceedings of the 12th ACM SIGIR Conference*, pp. 137–150, doi:10.1145/361219.361220.
- [26] Hinrich Schütze (1992): *Dimensions of Meaning*. In: *Proceedings of the 1992 ACM/IEEE conference on Supercomputing*, pp. 787–796, doi:10.1109/SUPERC.1992.236684.
- [27] Dan Sperber & Deirdre Wilson (1995): *Relevance: Communication and Cognition*, 2nd edition. Blackwell.
- [28] Peter D. Turney & Patrick Patel (2010): *From Frequency to Meaning: Vector Space Models of Semantics*. *Journal of Artificial Intelligence Research* 37, pp. 141–188, doi:10.1613/jair.2934.
- [29] Dominic Widdows (2004): *Geometry and Meaning*. CSLI Publications, Stanford, CA.