

**Stage Two Report:**  
**A Statistical Language Model for the**  
**Generation of Figurative Language**

by  
Stephen McGregor

An outline of a thesis to be submitted to the University of London  
for the degree of Doctor of Philosophy

Department of Electronic Engineering  
Queen Mary, University of London  
United Kingdom

September 2015

# Abstract

MIMO (Multiple Input Multiple Output) technology has been regarded as a practical approach to increase the wireless channel capacity and reliability.

Abstract goes here...

# Table of Contents

<b>Abstract</b>	<b>i</b>
<b>Table of Contents</b>	<b>1</b>
<b>1 A Model for Constructing Concepts on the Fly</b>	<b>2</b>
1.1 Dynamic Context Sensitivity . . . . .	4
1.2 Literal Dimensions of Co-Occurrence . . . . .	7
1.3 Interpretable Geometry . . . . .	11
1.4 A Computational Process . . . . .	16
<b>References</b>	<b>17</b>

## Chapter 1

# A Model for Constructing Concepts on the Fly

This chapter is concerned with a theoretical overview and a technical description of a novel distributional semantic model designed to map words into conceptually productive geometric relationships. Theoretically speaking, the model which will be described is based on some well travelled, if not entirely mainstream, ideas about the nature of language and mind:

1. Concepts are not stable; they are generated in response to unfolding situations in an unpredictable environment;
2. Lexical semantics are accordingly always underspecified, and always resolved in some environmental context;
3. There is no relationship of strict supervenience between language and concepts one way or the other, but instead a dynamic by which concepts invite communication, and language affords conceptualisation.

These ideas, which have been outlined throughout the previous chapter, are not the standard dogma of computational linguistics, which generally, and understandably, has modelled concepts as modular, portable entities, language as a likewise stable system of representations and rules, and the relationship between the two as one of source and contingent data. This structure-oriented approach to language and mind epitomises a project that (Petersen, 1992) has described as “finding the features, rules, and representations needed for turning rationalist philosophy into a research program,” (p. 89). As computer science and philosophy of mind increasingly interact at the vertex of cognitive modelling, culturally relative ideas about the connection between mental representations and linguistic symbols become incorporated into the very architecture of data structures, engendering a positive feedback loop by which the outputs of symbol manipulating information processing systems reinforce the premise that representations are stable entities which can be trafficked in the form of words according to the rules of a grammar.

I have no pretensions of instigating a paradigm shift in computer science: I do not claim that the methodology I will now describe represents a radical departure from the prevailing and highly productive approach to the computational modelling of language or knowledge. It is, rather, an attempt to build some consideration of the idea that minds are not populated by representations and that words are not static containers of meaning into using the existing computational paradigm. With this in mind, my model is predicated upon three interrelated desiderata, derived generally rather than in a one-for-one way from the points enumerated above:

1. The model should be dynamically sensitive to context;
2. The model should function in a way that is transparent and operationally interpretable;
3. The model should be situate words in spaces which are likewise geometrically interpretable.

In the following three sections, each of these requirements will in turn be analysed in the context of the underlying theoretical context. This analysis is performed with an eye towards the immediate project of designing a statistical model for mapping word-vectors to concepts, and each element of the profile of desirable properties will be explored with this in mind. Then finally, in a fourth section, the fundamental implementation of the model will be described in technical detail.

## 1.1 Dynamic Context Sensitivity

At the heart of the technical work described in this thesis is an insight which is broadly accepted by theoretical linguists and philosophers of language: word meaning is always contextually specified. This wisdom is built into the foundations of both formal semantics Montague1974 and pragmatics Grice1975, and is likewise taken into account in contemporary context-free approaches to syntax (?, ?). As evident from the implementations of conceptual models surveyed in the previous chapter, however, the computational approach has generally relied on the idea that concepts can, at some level of composition, be cast as essentially static representations. The tendency to treat concepts as self-contained ontological entities consisting of properties that are wholly or partly transferable is built into the fabric of the formal languages used to program computers, and indeed into the mechanisms of modular data processing systems with specific compartments for the storage and processing of data.<sup>1</sup>

With that said, the importance of context has certainly not been ignored by statistically oriented computer scientists. Indeed, (Baroni, Bernardi, & Zamparelli, 2014) make a case for vector space approaches to “disambiguation based on direct syntactic composition” (p. 254), arguing that the linear algebraic procedures used to compose words into mathematically interpretable phrases and sentences in these types of models

---

<sup>1</sup>It is perhaps not a coincidence that (?, ?) was a seminal figure in the description of both the logic of lattice theory that has motivated more recent developments in concept modelling such as formal concept analysis Wille and the modular architecture of memory and processing components that defined computers in the period before the advent of highly parallel processing.

result in a systemic contextualisation of words in their pragmatic communicative context. (Erk & Padó, 2008) outlines an approach that models words as sets of vectors including prototypical lexical representations capturing information about co-occurrence statistics and ancillary vectors representing *selectional preferences* [a la] Wilks1978 gleaned from an analysis of the syntactic roles each word plays in its composition with other words. These composite vector sets are then combined in order to consider the proper interpretation of multi-word constructs of lexically loose or ambiguous nouns and verbs. In subsequent work ErkEA2010, the same authors describe a model which selects *exemplar* word-vectors from, again, composites of vectors, in this case extracted from observations of specific compositional instances of the words being modelled. In the first instance, composition is the mechanism by which word meaning is selectively derived, while in the second instance observations of composition are the basis for constructing sets of representational candidates to be selected situationally.

The model presented in this thesis is motivated by a premise similar to the one explored by Erk and Padó: there should be some sort of selectional mechanism for choosing the way that a word relates to other words in context. I would like to push this agenda a even further, though. Following on ?’s Barsalou insight into the *haphazard* way in which concepts emerge situationally, and likewise ?’s Carston proposition regarding *ad hoc* conceptualisation, I propose that the mechanism for contextually mapping out conceptual relationships between representations of words should be as open ended as possible, ideally lending itself to the construction of novel conceptual relationships in the same way that the state space of possible word combinations offers an effectively infinite array of linguistic possibilities. In particular, I suggest that the ephemeral nature of concept formation can be modelled in terms of *perspectives* on the conceptual affordances of lexical relationships. Figure 1.1b Illustrates this point. From one point of view, *dog* and *cat* refer to exemplars of the conceptual category PETS, while *wolf* and *lion* are typical of the category PREDATORS. From a more taxonomically aligned point of view, though, *dog* and *wolf* group naturally in the CANINE category, while *cat* and *lion* clearly

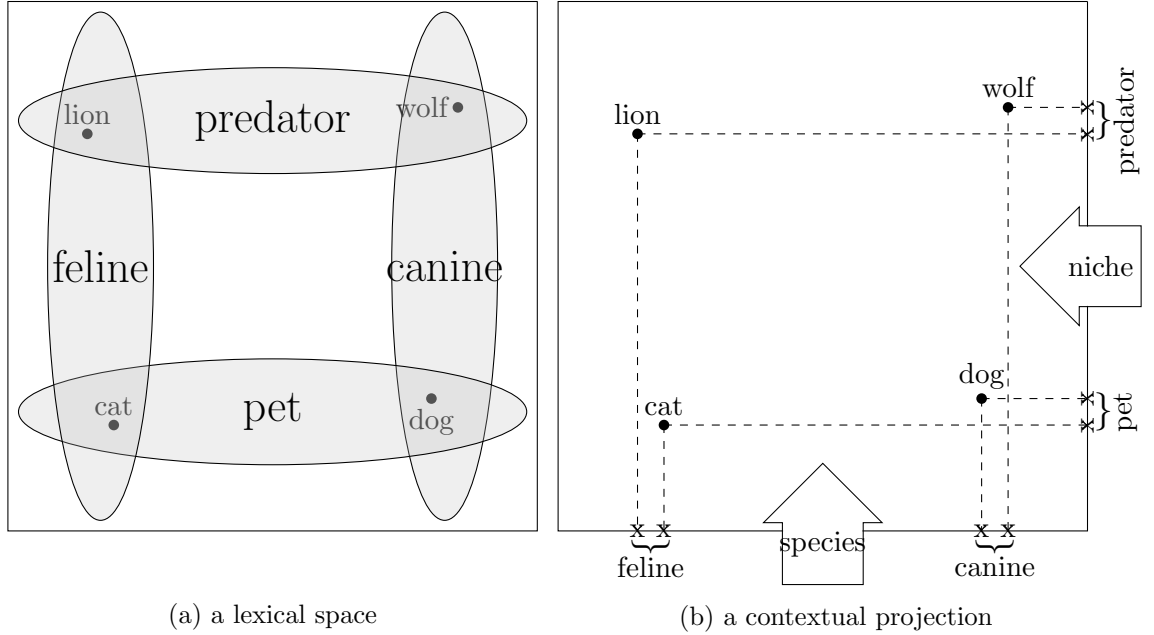


Figure 1.1: In the two-dimensional space depicted in (a), the conceptual vagary of four words maps to overlapping, elongated and indeterminate spaces. In (b), two different perspectives on the lexical space, represented by the arrows labelled *niche* and *species*, offer contextualised projections in one-dimensional clusters which remit conceptual clarity.

both belong to the category of FELINES.

Furthermore, the high dimensionality of vector space models of distributional semantics in particular should afford precisely these types of contextual vistas on potential relationships between words. Rather than depending on *a priori* disambiguation based on clustering or observations of context in the form of existing combinations of words, I propose that a technique for defining subspaces *in situ* will capture the momentary and situated way in which concepts come about in the course of a cognitive agent’s entanglement with the world. The way that relationships between words coalesce and then dissolve as we change our perspective on the space of this model is designed to reflect the way that concepts

This description of a process of perspective taking has thus far been quite high-level: that high dimensional spaces afford different vantage points on data is evident, but this begs the question of the actual mechanism for defining perspectives in such spaces. The



next section will focus on answering this question by outlining a method for

## 1.2 Literal Dimensions of Co-Occurrence

The model presented here is grounded within the paradigm of distributional semantics, which means that the conceptual geometries that it constructs are the product of observations of word co-occurrences in a large-scale corpus of textual data represented statistically. Two procedurally distinct methodological regimens have emerged from the recent study of distributional semantics. The first, and more established, approach involves tabulating word co-occurrence frequencies and then using some function over these to build up word-vector representations. With roots in the frequentist analysis described by (Manning & Humeau, 2008), recent research has typically involved matrix factorisation techniques presented as either (or both) an optimisation technique BullinariaEA2012 or a noise reducing mechanism KielaEA2014.<sup>2</sup> A more recent approach, which has received a great deal of attention with the increasing availability of large-scale data and the corresponding advent of complex neural network architectures, involves using non-linear regression techniques to iteratively learn word-vector representations in an online, step-wise traversal of a corpus BengioEA2003,CollobertEA2008,KalchbrennerEA2014. (Manning & Humeau, 2008) have described the former as *counting* and the latter as *predicting*, but it must be noted that both methods are very much grounded in observations about the co-occurrence characteristics of vocabulary words across large bodies of text.

Another important similarity between these two approaches is that they each in their own way move towards a representation of relationships between word-vectors which is to some extent optimally informative, and, by the same token, abstract. In the instance of neural network approaches, this is clearly the case due to the fundamental nature of the system: the dimensions of this variety of model exist as basically arbitrary handles

---

<sup>2</sup>(Manning & Humeau, 2008), (Manning & Humeau, 2008), and (Manning & Humeau, 2008) have all reported that dimensional reduction techniques including SVD, random indexing, and top frequency feature selection generally do not improve results on word similarity and composition tests, with some notable parameter specific exceptions.

for gradually adjusting the relative positions of vectors, slightly altering every dimension of each vector each time the corresponding word is observed in the corpus. And, as far as models based on explicit co-occurrence counts are concerned, the favoured technique tends to involve starting with a large, sparse space of raw co-occurrence statistics (frequencies, or, more typically, a mutual information type metric) and then factorising this matrix using a linear algebraic technique such as singular value decomposition. The result, in either case, is a space of vectors which exists just for the sake of placing words in a relationship where distance corresponds to a semantic property, consisting of dimensions which can only be interpreted in terms of the way that they allow the model to relate words, not in terms of their relationship to the underlying data. In fact, (Peters & Loper, 2008) have argued that recently developed neural networks approaches just exactly do recapitulate the process of matrix factorisation, and that a careful tuning of hyperparameters will generate commensurable results from either type of model.

A key feature of the model proposed in this thesis is that it maintains a space of highly sparse co-occurrence statistics, which, despite their anchoring in the relatively abstract realm of word positions in a digitised corpus, I will describe as *literal* in the sense that they can be interpreted as corresponding to actual relationships between words in the world. I don't wish to claim that there is scope for completely or even mainly recapitulating a nuanced conceptual model from the data available in a purely textual environment; to do so would be to move towards claims that intentionality can emerge from rule-based operations on symbols, and the problems with this have been explored by Searle's (1980) Chinese room argument and a subsequent generation of philosophers (Preston, 2002). But I would like to suggest that by building a base model that maintains the accessibility of unreduced co-occurrence information, we likewise maintain the ability to manipulate this base model extemporaneously, in reaction to the ongoing emergence of new contextual information. The idea is that such a base model would essentially represent the superset of all possible dimensions available for *ad hoc* selection in the course of a

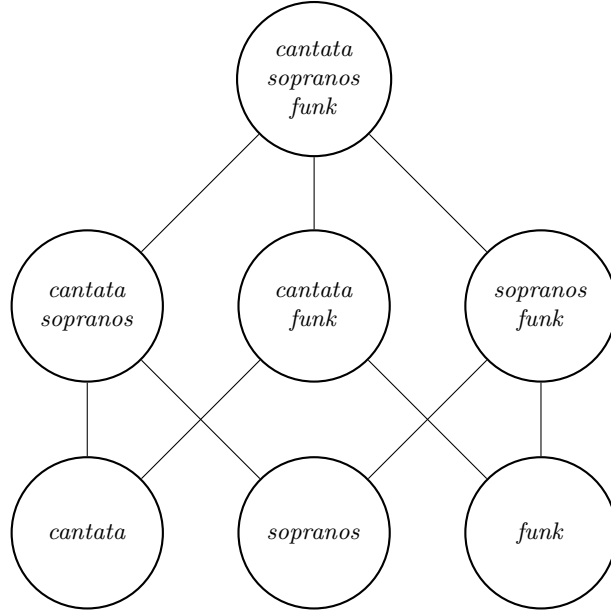


Figure 1.2: A Lattice of Three Dimensions

So the proper framework for describing the model to be examined in this thesis is not so much a single space of word-vectors as a Grassmannian lattice consisting of the power set of all possible combinations of the dimensions characterising the base space. At the top of this lattice – the *join* – sits a single  $k$ -dimensional space consisting of every available one of the  $k$  co-occurrence terms observed throughout the underlying corpus. At the bottom of the lattice – the *meet* – sit  $k$  different one-dimensional spaces, each space corresponding to a single co-occurrence term. If the meet is considered *layer* – 1 of the lattice, and the join is considered layer  $k$ , then any given interstitial *layer* –  $j$  consists of every possible combination of  $j$  dimensions of co-occurrence statistics. A diagram of a very simple example of one such model is presented in Figure 1.2, illustrating the possible subspaces projected from a vastly simplified model consisting of just three co-occurrence dimensions (these particular spaces will be explored in the next section, as illustrated in Figure 1.3).

An important distinction must be drawn, however, between the representation of my model as a lattice and the use of manifolds as an inferential mechanism. Formal concept analysis in particular has made a productive discipline out of applying lattice

type structures to conceptual modelling, using the semi-hierarchical properties of lattices to capture logical relationships of entailment Wille1982. That body of work takes as given that concepts are “the basic units of thought formed in dynamic processes within social and cultural environments,” [p. 2]Wille2005. (?, ?) offers a broad overview of how this approach might be pursued through corpus linguistic techniques, while (?, ?) and, more recently, (?, ?) have proposed statistical techniques using *feature inclusion* metrics to assess the potential entailment relationship between candidate words and corresponding concepts. The assumption inherent in this interesting work is that words are in some sense supervenient upon the concepts they denote, and that the statistical features of a language will by and large recapitulate the conceptual structure upon which it sits.

As (?, ?) has pointed out, however, it is problematic to assume that a spectrum of co-occurrence alone can indicate relationships of hyponymy and hypernymy. It stands to reason, for instance, that a word with a taxonomically specific referent such as *bulldog* should probably have a co-occurrence profile including words omitted from the corresponding profile of a word like *lifeform*, which has an ostensibly more general extent. Rimell has proposed a measure of change in *topic coherence* as word-vectors are combined algebraically in order to detect entailment relationships. This measuring is achieved specifically through a process of dimension-by-dimensions comparison between potentially related word-vectors, in particular the *vector negation* method described by (?, ?), combined with topic modelling techniques to analyse the coherence of features distilled by the selectional process.

The model proposed in this thesis adheres to the same principle of fine-grained cross-dimensional analysis described by Rimell. In addition to the practical issues raised by Rimell, my model is also designed to remain pointedly uncommitted to any idea that concepts are atomic or elementary to thought, or that language and concepts are involved in any kind of strictly hierarchical relationships. Instead, the model operates through an analytical traversal of a lattice of subspaces in search for a combination of dimensions that captures a conceptually *salient* profile of co-occurrence features. If a consequence of

this stance is that the model can't be understood in terms of nested ordered relationships, though, then the question of how conceptual relationships do emerge situationally from the model remains. The next section of this methodological overview will examine how the actual geometry of a projected subspace itself is expected to do this conceptual work.

### 1.3 Interpretable Geometry

It is important at this point to distinguish between two different modes of interpretability at play within the operation of the model I'm proposing. On the one hand, we have the mechanism for selecting subspaces described above: this mechanism requires a model composed of tractable dimensions of statistics that can be interpreted based on expectations generated from an analysis of some sort of contextually relevant information. Some specific mechanisms for this process will be discussed in the next section. Then on the other hand, once this selectional process has taken place, we find ourselves with a subset of dimensions defining a specific subspace. My claim is that, given the correct selectional criteria for performing this projection – this traversal of our lattice of vector spaces – we should be able to generate a subspace in which the projected word-vectors will be interpretable in terms of the actual geometric features of this subspace.

The idea of exploiting the geometry of a transformed space of word statistics is not new. Indeed, seminal work on latent semantic analysis (LSA) was motivated by precisely the insight that a singular value decomposition of a high-dimensional, sparse matrix of statistical data about word co-occurrences would result in a dense lower dimensional matrix in which dimensions characterise *latent semantics* rather than literal word co-occurrences DeerwesterEA1990. Thus the linear algebraic methodology of generating a lower dimensional matrix of optimally informative dimensions arguably transforms a space of specific co-occurrence events into a space of more general conceptual relationships. In fact, (?, ?) have subsequently argued that the dimensional reduction by way of factorisation itself might directly mirror cognitive conditioning, modelling the way that

the mind can “correctly infer indirect similarity relations only implicit in the temporal correlations of experience,” (p. 212).

Of course the dimensions of a factorised matrix are still not interpretable in themselves. They are, rather, an optimal abstraction of the underlying data, in which each dimension is maximally informative – and, accordingly, orthogonal – in comparison to the other dimensions. What we desire in a model, however, is a mechanism for actually interpreting directions and regions within a subspace projected by the model. This objective is motivated by (?, ?) insight into the inferential power of *conceptual spaces*: by building spaces in which the dimensions themselves correspond to *properties*, Gärdenfors has illustrated how features of points and regions within these spaces such as convexity and betweenness can be interpreted as corresponding to conceptual membership and can accordingly be used to reason about relationships between concepts. In more recent work, motivated by psycholinguist insight into the significance of the *intersubjectivity* by which language facilitates the mutual ascription of cognitive content between interlocutors, (?, ?) has proposed that semantics are derived from a communicative alignment of conceptual spaces.

A classic example of a Gärdenforsian conceptual space is the space of colors, which can be defined in terms of, for instance, hue, brightness, lightness, and colourfulness: any colour can be specified as a point corresponding to coordinates along each of these dimensions. Moreover, regions within the space of colours can be defined geometrically: the concept RED will correspond to a convex region within the space, and any point lying between two points known to be labelled *red* will likewise be considered *red*. (?, ?) has devised an experiment mapping linguistic descriptions to conceptual regions precisely within the domain of colours. Taking a large set of multi-lingual data regarding colour naming conventions, treating each of 330 different colours as an initially independent dimension, Jäger demonstrated how an extrapolation of optimally informational dimensions via a principle component analysis revealed clusterings of color names into convex

regions.<sup>3</sup>

Similarly motivated by Gärdenfors’s model of conceptual spaces, (?, ?) have built vectors of domain specific documents, associating word frequencies within documents with document labels. A multi-dimensional scaling procedure is then used to project these document-vectors into a Euclidean space in which the authors predict that properties such as *parallelness* and *betweenness* will correspond to conceptual relationships between documents. The authors demonstrate that geometry in their projected spaces does indeed afford conceptual interpretation: the word *bog* is found to be more or less between *heath* and *wetland*, for instance, and the vector for the film *Jurassic Park* lies in a direction associated with DINOSAURS and SPECIAL EFFECTS. This work is particularly notable in that ? appreciate the significance of projecting spaces which are interpretable in terms of Euclidean distances rather than simply the cosine similarity of vectors extending from the origin of a space: Euclidean metrics provide a platform for more nuanced considerations of the relationships between points.

The type of space exemplified by the research of Jäger and ? is moving towards being a conceptual space in the way that its geometry offers itself up to semantic interpretation, but importantly these remain static spaces comprised of abstract dimensions, albeit dimensions generated in order to optimise the interpretability of the spaces they delineate. The objective of my model is to emulate the geometric interpretability of these other spaces in an extemporaneous, contextually dynamic way. To illustrate this point, consider the two spaces illustrated in Figure 1.3 (taken from real co-occurrence data, as described in the next section, and based on the lattice of subspaces illustrated in Figure 1.2). Here co-occurrence statistics are used to define three different dimensions, from which two different two-dimensional subspaces are selected with word-vectors plotted into each subspace. In each subspace, a particular conceptual geometry emerges, oblique to the axes of each subspace but nonetheless indicating distinct conceptual regions in which

---

<sup>3</sup>The cross-cultural universality of colour naming conventions presented by (?, ?), which Jäger takes as a basis for his research, is controversial to say the least – see (?, ?) for an alternative point of view – but Jäger’s work remains a good example of a computational technique for extrapolating conceptual spaces from quantitative linguistic data.

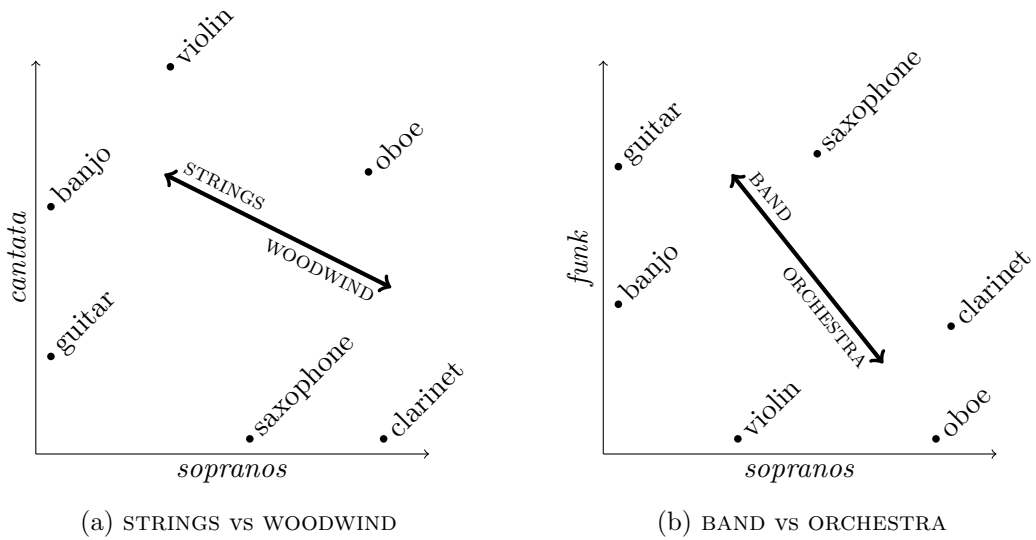


Figure 1.3: Based on real co-occurrence data, swapping one dimension in a two-dimensional subspace reveals two different conceptual geometries.

words cluster in an interpretable way.

The first thing to note about these spaces is the way that swapping a single dimension in a two dimensional subspace can have a significant impact on the conceptual affordances of the subspace’s geometry. The next thing to note is that the dimensions themselves are not especially interpretable: it might be surprising, for instance, that the word *cantata* has a higher propensity for co-occurrence with the word *banjo* than with the word *clarinet*, given that cantatas have traditionally included parts for the latter but not the former. An examination of the underlying data, extracted, as described in the next section, from English language Wikipedia, reveals that the term *cantata* has been adopted, perhaps somewhat figuratively, by some bluegrass musicians, and so co-occurrences with *banjo* are indeed observed. But rather than consider this usage anomalous or attempt some sort of *a priori* word sense disambiguation, I propose that we embrace the haphazardness of language and use it as a tool

The counterpoint remains that the dimensions in my model are also not in themselves interpretable, and so in this sense its spaces are not – at least not yet – Gärdenforsian. This is not surprising, and in fact it would be surprising if it turned out that in anything



other than the most specialised cases we could simply pick dimensions based on their labels and then expect co-occurrence statistics to play out in a conceptually coherent way, as this would contradict the Relevance Theoretic thesis that language in use is always significantly underspecified.

In the case of dimensions constituting a subspace projected by our model, we have a label which points us to a literal co-occurrence data as observed in a corpus.

The other variety of interpretability within this model happens on the level of the overall geometry of a conceptually oriented subspace.

and it is crucial to highlight the discrepancy between this geometric interpretability and the lower-level tractability that enable dimensional selection.

So, while a dimensional label  $c$  denoting a scalar for a vector  $\vec{w}$  can be interpreted as meaning that the word  $w$  co-occurs with the word  $c$  in a particular way relative to the overall frequency profiles of  $w$  and  $c$ , it does not necessarily mean anything more general about the relationship between any prospective concepts  $w$  and  $c$ . After all, concepts are in the end all constructed in an *ad hoc*, contextually specific way;

One important consequence of this distinction between dimensional and geometric interpretability is that, while the model can be properly understood as a lattice, it is not the case that this lattice can in turn be used to perform logical operations of entailment in the spirit of ?'s Wille formal concept analysis. The space at the top of the lattice – the ultimate join – does indeed contain all the information available to the model, but it can be thought of more as a muddle than as a comprehensive

## 1.4 A Computational Process

As is generally the case with data cleaning, these measures are prone to error

but one of the strengths of the subspace projection technique that my model uses is its

resilience to noise. So, for instance, misspellings will be categorised as highly anomalous co-occurrence dimensions and are therefore unlikely to be contextually selected (or, if they are regularly encountered enough to be contextually significant, there may well be

# References

(n.d.).

- Baroni, M., Bernardi, R., & Zamparelli, R. (2014). Frege in space: A program for compositional distributional semantics. *Linguistic Issues in Language Technology*, 9, 241–346.
- Baroni, M., & Lenci, A. (2010). Distributional memory: A general framework for distributional semantics. *Computational Linguistics*, 36(4).
- Baroni, M., & Zamparelli, R. (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* (p. 1183-1193).
- Bengio, Y., Ducharme, R., Vincent, P., & Jauvin, C. (2003). A neural probabilistic language model. *Journal of Machine Learning Research*, 3, 1137-1155.
- Birke, J., & Sarkar, A. (2006). A clustering approach for nearly unsupervised recognition of nonliteral language. In *Proceedings of eacl-06* (pp. 329–336).
- Black, M. (1955). Metaphor. In *Proceedings of the aristotelian society* (Vol. 55, p. 273-294).
- Black, M. (1977). More about metaphor. In A. Ortony (Ed.), *Metaphor and thought* (2nd ed., p. 19-41). Cambridge University Press.
- Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent Dirichlet allocation. *Journal of Machine Learning Research*, 3, 993-1022.
- Boden, M. A. (1990). *The creative mind: Myths and mechanisms*. London: Weidenfeld and Nicolson.

- Calvino, I. (1983/1985). *Mr. palomar*. New York: Harcourt Brace Jovanovich. (Translated by William Weaver)
- Chalmers, D. J. (1996). *The conscious mind*. Oxford University Press.
- Chomsky, N. (1959). A review of b. f. skinner’s verbal behavior. *Language*, 35(1), 26-58.
- Cimiano, P., Staab, S., & Tane, J. (2003). Automatic acquisition of taxonomies from text: Fca meets nlp. In *Proceedings of the ecml/pkdd workshop on adaptive text extraction and mining* (p. 10-17).
- Clark, S. (2015). Vector space models of lexical meaning. In S. Lappin & C. Fox (Eds.), *The handbook of contemporary semantic theory*. Wiley-Blackwell.
- Coecke, B., Sadrzadeh, M., & Clark, S. (2011). Mathematical foundations for a compositional distributed model of meaning. *Linguistic Analysis*, 36(1-4), 345–384.
- Collobert, R., & Weston, J. (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*.
- Colton, S., & Wiggins, G. (2012). Computational creativity: The final frontier? In *Proceedings of the 21st european conference on artificial intelligence*.
- Davidson, D. (1974). On the very idea of a conceptual scheme. In *Proceedings and addresses of the american philosophical association* (Vol. 47, p. 5-20).
- Deerwester, S., Dumais, S. T., Furnas, G. W., Landauer, T. K., & Harshman, R. (1990). Indexing by latent semantic analysis. *Journal for the American Society for Information Science*, 41(6), 391-407.
- Dennett, D. (1996). *Kind of minds*. London: Weidenfeld & Nicolson.
- Dummett, M. (1981). *Frege: Philosophy of language* (2nd ed.). London: Duckworth.
- Dunn, J. (2013). Evaluating the premises and results of four metaphor identification systems. In A. Gelbukh (Ed.), *Computational linguistics and intelligent text processing: 14th international conference, cicling 2013, samos, greece, march 24-30, 2013, proceedings, part i* (pp. 471–486). Springer Berlin Heidelberg. Available from [http://dx.doi.org/10.1007/978-3-642-37247-6\\_38](http://dx.doi.org/10.1007/978-3-642-37247-6_38)

- Eco, U. (1976). *A theory of semiotics*. Bloomington: Indiana University Press.
- Erk, K., & Padó, S. (2008). A structured vector space model for word meaning in context. In *Proceedings of the conference on empirical methods in natural language processing* (pp. 897–906). Available from <http://dl.acm.org/citation.cfm?id=1613715.1613831>
- Erk, K., & Padó, S. (2010). Exemplar-based models for word meaning in context. In *Proceedings of the acl 2010 conference short papers* (pp. 92–97). Available from <http://dl.acm.org/citation.cfm?id=1858842.1858859>
- Evans, V. (2009). *How words mean: Lexical concepts, cognitive models, and meaning construction*. Oxford University Press.
- Fass, D. (1991). Met\*: A method for discriminating metonymy and metaphor by computer. *Computational Linguistics*, 17(1), 49–90. Available from <http://dl.acm.org/citation.cfm?id=971738.971741>
- Fillmore, C. J. (1976). Frame semantics and the nature of language. *Annals of the New York Academy of Sciences*, 280(1), 20–32.
- Firth, J. R. (1959). A synopsis of linguistic theory, 1930-55. In F. R. Palmer (Ed.), *Selected papers of j. r. firth 1952-59*.
- Fodor, J. A. (1975). *The language of thought*. Cambridge, MA: Harvard University Press.
- Fodor, J. A. (2008). *Lot 2: The language of thought revised*. Oxford University Press.
- Fox, C., & Lappin, S. (2005). *Foundations of intensional semantics*. Oxford: Blackwell Publishing.
- Gargett, A., & Barnden, J. (2013). Gen-meta: Generating metaphors using a combination of ai reasoning and corpus-based modeling of formulaic expressions. In *Proceedings of taai 2013*.
- Gibbs, R. W., Jr. (1994). *The poetics of mind*. Cambridge University Press.
- Grice, H. P. (1975). Logic and conversation. In P. Cole & J. L. Morgan (Eds.), *Syntax and semantics volume 3: Speech acts* (pp. 41–58). New York: Academic Press.

- Gutiérrez, E. D., Shutova, E., Marghetis, T., & Bergen, B. K. (2016). Literal and metaphorical senses in compositional distributional semantic models. In *Proceedings of the 54th annual meeting of the association for computational linguistics*.
- Harris, Z. (1954). Distributional structure. *Word*, 10(23), 146–162.
- Indurkha, B. (1997). Metaphor as change of representation: an artificial intelligence perspective. *Journal of Experimental & Theoretical Artificial Intelligence*, 9(1), 1-36. Available from <http://dx.doi.org/10.1080/09528139714721110.1080/095281397147211>
- Jordanous, A. K. (2012). *Evaluating computational creativity: A standardised procedure for evaluating creative systems and its application* Unpublished doctoral dissertation. University of Sussex.
- Kintsch, W. (2000). Metaphor comprehension: A computational theory. *Psychonomic Bulletin & Review*, 7(2), 257–266. Available from <http://dx.doi.org/10.3758/BF03212981> 10.3758/BF03212981
- Koestler, A. (1964). *The act of creation*. London: Hutchinson.
- Lakoff, G., & Johnson, M. (1980). *Metaphors we live by*. University of Chicago Press.
- Langacker, R. W. (1987). *Foundations of cognitive grammar: Theoretical prerequisites*. Stanford, CA: Stanford University Press.
- Lee, M. G., & Barnden, J. A. (2001). Reasoning about mixed metaphors within an implemented artificial intelligence system. *Metaphor and Symbol*, 16(1-2), 29-42.
- Levy, O., Remus, S., Biemann, C., & Dagan, I. (2015). Do supervised distributional methods really learn lexical inference relations? In *Proceedings of the 2015 conference of the north american chapter of the association for computational linguistics*.
- McGregor, S., Wiggins, G., & Purver, M. (2014). Computational creativity: A philosophical approach, and an approach to philosophy. In *Proceedings of the fifth international conference on computational creativity*.
- Mikolov, T., tau Yih, W., & Zweig, G. (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* (p. 246-251).
- Mitchell, J., & Lapata, M. (2010). Composition in distributional models of semantics. *Cognitive Science*, 34(8), 1388-1439.
- Narayanan, S. (1999). Moving right along: A computational model of metaphoric reasoning about events. In *Proceedings of the sixteenth national conference on artificial intelligence and eleventh conference on innovative applications of artificial intelligence* (pp. 121-127). Available from <http://www.aaai.org/Library/AAAI/1999/aaai99-018.php>
- Ortony, A. (1975). Why metaphors are necessary and not just nice. *Educational Theory*, 25(1), 45-53.
- Ortony, A. (Ed.). (1993). *Metaphor and thought* (2nd ed.). Cambridge University Press.
- Pantel, P. (2005). Inducing ontological co-occurrence vectors. In *Proceedings of the 43rd annual meeting on association for computational linguistics* (pp. 125-132). Stroudsburg, PA, USA: Association for Computational Linguistics.
- Peirce, C. S. (1931). *Collected papers of charles sanders peirce*. Harvard University Press. (edited by Charles Hartshorne and Paul Weiss)
- Pennington, J., Socher, R., & Manning, C. D. (2014). Glove: Global vectors for word representation. In *Conference on empirical methods in natural language processing*.
- Pietarinen, A.-V. (2004). Grice in the wake of peirce. *Pragmatics & Cognition*, 12(2), 295-315.
- Pinker, S. (1994). *The language instinct: How the mind creates language*. William Morrow.
- Pustejovsky, J., Rumshisky, A., Plotnick, A., Jezek, E., Batiukova, O., & Quochi, V. (2010). Semeval-2010 task 7: Argument selection and coercion. In *Proceedings of the 5th international workshop on semantic evaluation* (pp. 27-32). Available from <http://dl.acm.org/citation.cfm?id=1859664.1859669>
- Ritchie, G. (2007). Some empirical criteria for attributing creativity to a computer program. *Minds and Machines*, 17(1), 67-99.
- Russell, B. (1905). On denoting. *Mind*, 14(56), 479-493.

- Russell, B. (1931). *The scientific outlook*. London: George Allen and Unwin.
- Santus, E., Lenci, A., Lu, Q., & im Walde, S. S. (n.d.). Chasing hypernyms in vector spaces with entropy. In *Proceedings of the 14th conference of the european chapter of the association for computational linguistics*.
- Searle, J. R. (1979). Metaphor. In A. Ortony (Ed.), *Metaphor and thought*. Cambridge University Press.
- Shutova, E. (n.d.). Metaphor identification as interpretation. In *Proceedings of \*sem 2013*.
- Shutova, E. (2010). Models of metaphor in nlp. In *Proceedings of the 48th annual meeting of the association for computational linguistics* (pp. 688–697).
- Shutova, E. (2015). Design and evaluation of metaphor processing systems. *Computational Linguistics*, 41(4), 579–623.
- Shutova, E., Teufel, S., & Korhonen, A. (2012). Statistical metaphor processing. *Computational Linguistics*, 39(2), 301–353.
- Skinner, B. F. (1957). *Verbal behavior*. Acton, MA: Copley Publishing Group.
- Snow, R., Jurafsky, D., & Ng, A. Y. (2006). Semantic taxonomy induction from heterogeneous evidence. In *Proceedings of the 21st international conference on computational linguistics and the 44th annual meeting of the association for computational linguistics*.
- Socher, R., Chen, D., Manning, C. D., & Ng, A. Y. (2013). Reasoning with neural tensor networks for knowledge base completion. *Advances in Neural Information Processing Systems*, 26.
- Sweetser, E. (1990). *From etymology to pragmatics: Metaphor and cultural aspects of semantic structure*. Cambridge University Press.
- Thomas, M. S. C., & Mareschal, D. (1999). Metaphor as categorisation: A connectionist implementation. In *Proceedings of the aisb '99 symposium on metaphor, artificial intelligence, and cognition*. University of Edinburgh.
- Tsvetkov, Y., Boytsov, L., Gershman, A., Nyberg, E., & Dyer, C. (2014). Metaphor detection with cross-lingual model transfer. In *ACL (1)* (pp. 248–258). The Asso-



- ciation for Computer Linguistics.
- Turney, P. D., Neuman, Y., Assaf, D., & Cohen, Y. (2011). Literal and metaphorical sense identification through concrete and abstract context. In *Proceedings of the 2011 conference on empirical methods in natural language processing* (pp. 680–690).
- Utsumi, A. (2011). Computational exploration of metaphor comprehension processes using a semantic space model. *Cognitive Science*, 35(2), 251–296.
- van Genabith, J. (2001). Metaphors, logic and type theory. *Metaphor and Symbol*, 16(1-2), 23-57.
- Veale, T. (2016, June). Round up the usual suspects: Knowledge-based metaphor generation. In *Proceedings of the fourth workshop on metaphor in nlp* (pp. 34–41). San Diego, California: Association for Computational Linguistics. Available from <http://www.aclweb.org/anthology/W16-1105>
- Veale, T., & Hao, Y. (2007). Comprehending and generating apt metaphors: A web-driven, case-based approach to figurative language. *AAAI*, 1471-1476.
- Veale, T., Valitutti, A., & Li, G. (2015). Twitter: The best of bot worlds for automated wit. In N. Streitz & P. Markopoulos (Eds.), *Distributed, ambient, and pervasive interactions: Third international conference, dapi* (pp. 689–699). Springer International Publishing. Available from [http://dx.doi.org/10.1007/978-3-319-20804-6\\_63](http://dx.doi.org/10.1007/978-3-319-20804-6_63) 10.1007/978-3-319-20804-6\_63
- White, G. (2009). On scholarship. In T. Bartscherer & R. Coover (Eds.), *Switching codes: Thinking through digital technology in the humanities and the arts*. University of Chicago Press.
- Wiggins, G. A. (2006). Searching for computational creativity. *New Generation Computing*, 24, 209-222.
- Wilks, Y. (1978). Making preferences more active. *Artificial Intelligence*, 11(3), 197–223.

Wittgenstein, L. (1922). *Tractatus logico-philosophicus*. London: Kegan Paul, Trench, Trubner & Co, Ltd.

Wittgenstein, L. (1953/1967). *Philosophical investigations* (3rd ed.). Oxford: Basil Blackwell. (trans. G. E. M. Anscombe)