Metaphor Generation through Context Sensitive Distributional Semantics

Stephen McGregor, Matthew Purver, and Geraint Wiggins

Abstract: In this paper, we outline a preliminary methodology for generating metaphor based on contextual projections of representations built up through a statistical analysis of a large scale linguistic corpus. These projections involve defining subspaces of co-occurrence statistics in which we show that metaphors can be modelled as mappings between congruent regions of semantic representations. We offer this methodology as an empirical implementation pointing towards a resolution of theoretical stances, at times incompatible, construing metaphor as on the one hand an artefact of underlying cognitive processes and on the other hand a product of the environmentally situated generation of ephemeral conceptual schemes.

1 Introduction

In this chapter we discuss an idea for a novel computational methodology for metaphor generation. This methodology is inspired by a foray into the rich theory that has sought to address questions of how both how and why humans so ubiquitously resort to metaphor in their communications with one another. The account that we embrace here is, largely, pragmatic in nature, in that we view metaphor as a mechanism for using words to accomplish communicative objectives in a particular complex environmental setting, rather than as a computation of ways in which elements of a semantic representation can be projected across conceptual domains. Our aim is not to reject the idea of a more representational approach to metaphor, but we would like to consider what might be involved in a computational approach to representing metaphor generation as something that happens in a situated way, treating language as an artefact of the environment as well as of the mind. A consequence of our accommodation of this stance on metaphor

This is the author accepted manuscript of McGregor, Purver, Wiggins (2020). Metaphor Generation through Context Sensitive Distributional Semantics.
In *Producing Figurative Expression: Theoretical, experimental and practical perspectives* pp.419-448, DOI 10.1075/ftl.10.15mcg.

is that it becomes very difficult to consider how to appropriately evaluated the success of a novel metaphor: while we offer some thoughts on how this might happen, what we present here is an idea rather than a quantitative research effort.

Our proposed methodology is rooted in *distributional semantics*, a paradigm, well-known amongst computational linguists, for building up word representations from statistical data about observations of word use in such a way that the meaning of words is in a certain sense built into the representation itself, and into the way that representations of different words interact with one another. This technique has already been productively applied to a variety of natural language processing tasks. The essential novel feature of our methodology is that it offers a mechanism for the situational contextualisation of representations, allowing for the projection of extemporaneous semantic subspaces, representing, in the parlance of relevance theory, *ad hoc* conceptualisations (see Section 2), in which word representations are expected to relate to one another and to the space itself in ways that pertain to a specific conceptual scheme. We take it as given that the idea of representing word meaning in a context specific way is desirable, and indeed our methodology has already been applied productively to metaphor classification (Agres et al., 2016) as well as the task of classifying the related non-literal phenomenon of semantic type coercion (McGregor et al., 2017).

Beyond offering an outline of a computational technique for generating metaphor, this work has been conceived as an empirical mechanism for exploring a theoretical issue in the study of metaphor. In particular we would like to consider the productive dichotomy between cognitive linguistic approaches to metaphor, which have conceived of metaphorical language as a manifestation of underlying metaphorical cognitive processes, and pragmatic accounts of metaphor, which, particularly under the banner of relevance theory, have considered figurative language as the deep end of a gradient of loose language use corresponding to the situational way in which concepts are formed by linguistic agents. There has already been productive dialogue between proponents of each perspective, which we will briefly survey below. We think of our methodology as an instantiation of a theoretical stance which seeks to incorporate insight from both theoretical camps. We would especially like to present our work as a preliminary account of how a statistical technique can be applied to the pragmatic account of metaphor, something that has proven challenging for computational methodologies.

It is worth mentioning from the outset that our notion of *context* is itself based on a pragmatic interpretation of how meaning comes about in the process of a linguistic agent's interaction with other agents in the world. As such, we will talk about context as the overall situation of being in the world, including various non-lingusitic components of the experience of existence as a cognitive entity. In this regard, our application of *context* will be considerably more expansive than the meaning of the word sometimes encountered in computational linguistics, where *context* is taken to indicate only the textual surroundings in which a particular unit of analysis, such as a word, is observed—though this latter notion is an essential component of the methodology we will describe, and we will refer to this as *co-occurrence*. In practice, the degree to which a computational process based on an anaysis of realtively abstract data can capture this more fully embedded notion of context is one of the fundamental questions behind the methodology we would like to explore.

In particular, our methodology generates context specific subspaces based on input in the form of sets of words: if there is a semantic coherency to such a set of words, then our expectation is that the projection of word representations, in the form of vectors, into this subspace will be endowed with some of the conceptual nuance corresponding to that semantic context. Our methodology then facilitates the exploration of the geometric relationships between words in these subspaces. We propose that metaphor in particular can be understood in terms of mappings between conceptual domains within subspaces, defined as arrangements of semantically associated word-vectors. This commitment lends itself both to visualisation (in three dimensional spaces) and to straightforward computation of new metaphors (in arbitrarily high dimensional spaces). Importantly, our geometries can be interpreted in terms of tendencies of word use observed in an underlying corpus, meaning that the operation of our methodology can be pushed back out of the computations themselves and into expectations about how tendencies in language use relate to the construction of metaphor.

The work described here, while it includes some concrete examples of how our methodology can potentially operate, is an initial description of a general framework: we are not yet at the point where we can simply input a few words into a terminal and expect a well-formed, functional metaphor to be output. As such, we do not claim to present a comprehensive empirical model of context sensitive metaphor generation here; rather, we outline some of the theoretical issues at stake, and propose a data driven, statistically grounded approach that we suggest has the potential to address these issues. In fact, our claim is that the high dimensional representational spaces associated with distributional semantics, more than just presenting a possible technique for metaphor generation, offer certain properties that are particularly appealing when considering the way that metaphor emerges contextually in the process of lingusitic agents' intereaction with one another in an unpredictable world.

What we can do is work backwards from instances of known metaphors, establish that there are configurations of our technique in which these metaphors can be generated, and then consider the ways in which we might rework this reversed procedure in order to produce new metaphoric mappings based on input consisting of words indicating a target domain and words indicating the context in which a metaphor might arise. We will offer a preliminary consideration of what some of the next computational steps in the implied research project might look like. In the end, the evaluation of computer generated metaphor, particularly creative metaphor, is in itself a significant challenge. While our work is not yet at the stage where a meaningful evaluation is possible, we suggest that an evaluator would need to in some sense take into account the context in which a metaphor is generated in order to determine the quality of the output.

Finally, we will argue that the directness of our geometric methodology offers a way towards an ecologically situated model of language, in which metaphor comes about as linguistic agents grasp for whatever semantic objects they can find for communicating about situations as they unfold in an unpredictable environment. By this theory, opportunities for meaning-making are directly perceived in the affordances of the linguistic components available to an agent *in situ*, and the application of lexicalised terms in surprising semantic contexts is an expected outcome of this scenario. Language then becomes, as Clark (2006) puts it, " a kind of self-constructed cognitive niche: a persisting but never stationary material scaffolding whose crucial role in promoting thought and reason remains surprisingly poorly understood," (p. 370).

2 Language in Minds, Minds in the World

At the root of much of the theoretical and psycholinguistic work that has been done on metaphor over the past several decades, and indeed some of the more recent computational advances in the field (see Section 3) is the idea that "we systematically use inference patterns from one conceptual domain to reason about another conceptual domain," (Lakoff and Johnson, 2003, p. 246). At least one of the appealing features of a model of metaphor rooted in conceptual structures and isomorphisms between these structures is that it provides a mechanism for a dynamic interplay between the semantic commitments of metaphor in language and the corresponding features of cognition (Gibbs Jr., 1994). The commitment to a close link between metaphor as it arises in language and corresponding cognitive processes has allowed for productive mappings between domains of human experience in terms of the application of metaphoric language: through the study of metaphors projecting from the domain of the body to that of the mind, for instance (Sweetser, 1990), or projecting from orientation in space to conception of time (Fauconnier and Turner, 2008).

An open question, though, regards the role of concepts themselves in both language and cognition. It would be convenient for the conceptual approach to metaphor if concepts could be modelled as well-defined composite structures, characterised by components that afford mappings between their parts. Somewhat contrarily, however, Barsalou (1992, 1993) has described the *ad hoc* nature of conceptualisation as it occurs in the complex cognitive environment of a linguistic agent: Barsalou examines the way in which concepts emerge as recursive concatenations of *frames* that come about in the course of goal-directed action in a situation in the world. By Barsalou's account, frames provide a structure to a concept, accommodating representational roles for *attributes* and *values*. Importantly, though, these structures come about *contextually*: a cognitive agent constructs a frame in response to a specific situation encountered in its environment, based to some extent on prototypical attributes and values. The *ad hoc* nature of these concpetual structures contributes in a fundamental way to the flexibility inherent in language in use.

Applying this contextually situated cognitive framework specifically to metaphor, Carston (2002) points out that there are a variety of cases (paradigmatically “Bill is a bulldozer”) where there is no clearly explicit transfer of properties from the domain of the source to the domain of the target; rather, the *ad hoc* concept BULLDOZER* seems to take on, in the course of its context specific generation, properties that are peculiar to the domain of human action and affect. Without claiming to have landed on a definitive solution to the problem of mappings between evidently disparate domains, one path towards an answer suggested by Carston is that there could be mappings discovered specifically by way of a kind of congruency between the structures of domains, correlations which would presumably not necessarily fall back on any actual conceptual overlap between the components of the structures (see Gentner, 1983, for a foundational exploration of *structure mapping*).

A productive dialogue has subsequently evolved between the cognitive and pragmatic camps. So, for instance, Gibbs Jr. and Tendahl (2006) propose that the mappings between disjoint stable and *ad hoc* conceptual domains – the mapping from, for example, MACHINE to HUMAN – are in fact encoded as encyclopaedic conceptual metaphors that offer a conduit that might be hidden from an analysis of the more specific conceptual properties at play in a particular expression. Carston (2012) responds by acknowledging that there may be a systematisation at work in the metaphoric linking of certain broad conceptual domains, but maintains that the underlying cognitive mechanisms behind these links involve patterns of abstract thought rather than metaphoric extrapolation from relatively concrete domains, with metaphor itself emerging as a communicative rather than cognitive mechanism. The upshot of this discourse is twofold: on the one hand, there is clearly much to be gained from the interaction between the cognitive stance on the fundamental metaphoricity of thought and the pragmatic framework of situated conceptualisation (and indeed, much has been gained); on the other hand, there is still an important unanswered question regarding the conceptual foundations of the transmission of information via metaphor.

Indications of a middle way might be found in the controversial theoretical position taken by Davidson (1978), who famously suggests that "metaphors mean what the words, in their most literal interpretation, mean, and nothing more," (p. 245). Davidson's point is not that metaphors are in some sense secondary to the propositional content of the language conveying them; rather, his goal is to make a distinction between the philosophically risky notion of *meaning* as it relates to lexical representations – words – and a separate notion of *what a metaphor makes us notice* which is associated with the effects of metaphoric utterances as opposed to the interpretation of the semantic content of a metaphor. Ultimately, Davidson is denying that metaphors, in their capacities as insight bearing functionaries, have any propositional content at all. As Reimer (2001) puts it in her analysis of Davidson's stance, "the goal of the metaphor-maker is not to get the hearer to see that something is the case, to grasp some deep and subtle truth, but to see something in a certain way, and seeing something in a certain way is simply not the sort of thing that can be given literal expression," (p. 150).

The Davidsonian approach to metaphor presents a particular challenge to computational modelling: how are symbol manipulating systems to get at the essence of a metaphor if it specifically lies beyond the boundaries of the type of propositional content of structured lexical semantic representations that computers are capable of handling? The methodology described in this paper has been constructed specifically to address this theoretically appealing, technically challenging dilemma. This novel approach is built upon a semantic model that construes lexical representations as points in high dimensional spaces, and incorporates the insight that in such spaces, context specific semantic relationships can be understood as lower dimensional perspectives on an underlying representational structure. Because these representations exist alongside one another in an abstract and relatively unstructured statistical space, the propositional value of the geometric relationships to one another is at best ambiguous. Rather than constructing metaphors as mappings between highly structured conceptual representations, then, our approach models metaphor in terms of transformations across congruent geometric semantic structures in subspaces which are themselves projected from a base lexical space generated from an analysis of large-scale textual data. These transformations within projections arise fleetingly, as the ephemera of *ad hoc* conceptualisations that occur in the situation of a cognitive context.

As such, the inauguration of this computational approach to metaphor is necessarily generative rather than interpretive. This is because the methodology does not presume that the production of a metaphor is underpinned by a propositional interpretation; in fact, the strength and theoretical thrust of the approach is that metaphors are discovered as geometric elements of semantic spaces, affording the opportunity for communication without necessarily explicating any communicated content. In this regard, we propose that our approach, while incorporating some of the cognitive linguistic insight which has been productive in existing computational approaches to metaphor, simultaneously opens itself up to a pragmatic account of concepts as situated and temporary constructs, and indeed to Davidson's outright rejection of structured cognitive content as a feature of metaphor. Ultimately, we believe that our approach has the potential to accommodate a linguistic adaptation of the ecological psychology of Gibson (1979), who describes cognition in terms of the direct perception of opportunities for action in an environment: just as a cognitive agent perceives objects in terms of the goal-oriented actions that they afford, a linguistic agent perceives language generation in terms of opportunities for communication, and so metaphors emerge as the product of such an agent taking up lexical entities and using them as best fits some communicative requirement.

3 Computational Approaches to Metaphor

As Shutova (2015) has noted in her comprehensive survey of the field, the *selectional preference* approach of Wilks (1978) has been fundamental to a certain mode of computational metaphor modelling. By this account, metaphor involves the projection of the conceptual attributes associated with one conceptual structure onto another: words are associated with structured representations, and metaphor involves the resolution of instances of *preference breaking* where the attributional expectations of a semantic frame are not satisfied by the supplied term. A type theoretical application constructed along similar lines is described by van Genabith (1999), who makes use of the observation that metaphor can, at least sometimes, be treated as a kind of condensed mode of simile. From a theoretical standpoint, this approach lines up somewhat with Searle's (1979) view on metaphor, which holds that metaphor interpretation involves an identification of a semantic dislocation followed by a search for an inferential interpretation based upon beliefs about a metaphoric source that somehow apply to its target. There is, for that matter, a correspondence with the pragmatic account of Grice (1975), which casts metaphor in terms of the implicatures signalled through the activation of communicative norms (see Reimer, 2013, for a recent exploration of Gricean pragmatics in light of subsequent research on metaphor).

Elsewhere, however, cognitive linguistic approaches to metaphor, and in particular the conceptual metaphor model of Lakoff and Johnson (2003), have provided fertile theoretical material for computational linguistic applications. So, for instance, the ATT-Meta system of Barnden and Lee (1999) specifically targets "useful, common sense reasoning based on metaphor" (p. 28) through a system for developing novel hypotheses about conceptual relationships across elements of established metaphorical domain mappings. And the CorMet system of Mason (2004) extrapolates mappings of a dataset of conceptual metaphors from an analysis of the WordNet lexical taxonomy. Kintsch's (2000) work on metaphor comprehension is likewise influenced by the idea that metaphors emerge from mapping between conceptual domains, rather than as computations arising from resolving violations of categorical inclusion.

Conceptual metaphor theory has been particularly influential on relatively recent developments in data-driven, statistical approaches to metaphor that take advantage of a combination of growing computational power and increasing access to large-scale digitised corpora. In this spirit, Shutova et al. (2013) use word co-occurrence patterns to discover conceptual metaphoric mappings as correspondences in usage between concrete and abstract domains. Similarly, Tsvetkov et al. (2014) train a model to classify metaphoric expressions based on statistical characteristics of the words involved in phrases, using co-occurrence profiles as the units of analysis in predicting, for instance, the abstractness of particular concepts.

These last cases of data-driven systems for metaphor classification and interpretation are examples of a general trend by which statistical approaches to metaphor generally operate, directly or indirectly, in the vicinity of the distributional semantic paradigm. Dating back to the theoretical research of Harris (1954) and the applied work of Salton et al. (1975) and Schütze (1992), among others, this approach to natural language processing involves quantifying observations about the way in which words tend to co-occur with one another over the course of large-scale textual corpora. Working off the basic intuition that, as Pantel (2005) puts it, "words that occur in the same contexts tend to have similar meaning," (p. 126), the objective is to build up informationally rich lexical representations that facilitate quantitative analysis of semantic relationships. By representing words as vectors in a space defined in terms of axes of co-occurrence tendencies, the geometry of the space itself is endowed with semantic interpretability: so, through either the tabulation of collocations or the gradual adjustment of representations across iterative traversals of of a corpus (Mikolov et al., 2013). Clark (2015) offers a comprehensive overview of the field, and our own application of distributional semantic modelling will be discussed in the following section.

The conceptual metaphor model is particularly conducive to quantitative operations on computationally tractable representations. To offer a little background here, Coecke et al. (2011), among others, have described a compositional language model that represents words as linear-algebraic structures, in particular combinations of vectors and higher order tensors.[1] The upshot of this data-intensive approach to modelling meaning is that, by assigning different representational structures to different grammatical classes, a model can describe the composition of words into sentences in terms of products between mathematical entities. So, for instance, Baroni and Zamparelli (2010) have shown that, by representing nouns as vectors and adjectives as matrices,

[1]Those authors have in particular sought to cash their compositional approach out in terms of the quantification of propositional statements; see also Grefenstette et al. (2014).

a noun-adjective phrase can be treated as a transformation where the operation of the adjective-matrix maps the noun-vector into the proximity of other words and terms conceptually related to the phrase being modelled.

Gutiérrez et al. (2016) offer an exemplary implementation of the way in which this space-transforming approach can be applied to the modelling of metaphor: following on the work in word sense disambiguation of Kartsaklis and Sadrzadeh (2013), they construct separate matrices based on observations of metaphoric and literal usage of a give adjective, reasoning that the model should then learn to map adjective-noun phrases according to the respective interpretations of the adjectives in question. So, for instance, *brilliant girl* should be mapped to a region of a semantic space occupied by other phrases such as *clever boy* and *intelligent child*, while *brilliant diamond* should be mapped into the vicinity of things like *shining emerald* and *bright stone*. They apply their model to the task of classifying metaphor by evaluating the proximity of the transformation of metaphoric versus literal representations of adjectives to the distributional profiles observed through treating the phrase itself as a unit of analysis, anticipating that the linear algebraic composition of $\textit{brilliant}_{\mathrm{MET}}$ and *boy* should be closer to *brilliant boy* than that of $\textit{brilliant}_{\mathrm{LIT}}$ and *boy*. The authors generalise their approach within the theoretical context of conceptual metaphor by learning broader mappings between conceptual domains that can be applied to compositions of noun-vectors with literal adjective-matrices to discover figurative transformations.

A nice thing about the application of compositional operations to distributional semantic representations is that it facilitates a model of metaphors as conceptual mappings. A more problematic aspect of this approach is that an already established propositional interpretation has to be built into the transformative representations based on something explicitly observed in the underlying data. Therefore, a given transformation depends either on the observation of existing instances of a particular component of a conceptual metaphor, and a corresponding *de facto* lexicalisation of a metaphorical sense of a word, or else on the prior configuration of a network of conceptual mappings. We would like to propose a distributional technique that makes the most of the transformational characteristics of linear algebraic representations, but that also accommodates the *ad hoc*, context specific way in which metaphor comes about, allowing for the open ended association of conceptual domains. We would also like to remain open to Davidson's (1978) suggestion that "there are no unsuccessful metaphors," (p. 245), or at least that the situations in which language might be engaged are so various that we should be prepared for effectively any unlikely linguistic composition to be applied.

A metaphor is, as such, in our approach, first discovered as an opportunity for contextualised communication, and only subsequently interpreted in terms of any cross-conceptual alignments that can be extrapolated from the metaphor—if in fact any such interpretation ever occurs, since it should be at least feasible that a metaphor can function as an effective communicative device without ever arriving at an explication of propositional content. This means that this first pass at a framework for *ad hoc* metaphor will be about generation rather than classification, and so it is appropriate to briefly consider existing work on this particular aspect of the computational modelling of metaphor, as well. Veale and Hao (2007) mine the web using syntagmatic heuristics indicative of attribute-projecting language in order to build up a knowledge base that, in conjunction with access to a lexical taxonomy indicating entailment relationships, can be used to generate novel metaphoric adjective-noun combinations: so conceptual mappings are learned from clues discovered in linguistic surface forms. Gargett and Barnden (2013), on the other hand, have described the Gen-Meta system for metaphor generation, based on the above mentioned ATT-Meta approach to metaphor interpretation and applying the logic that an interpretive process can potentially be construed as a generative process in reverse, effectively applying broad but established mappings between domains to discover more particular novel instances of metaphoric associations.

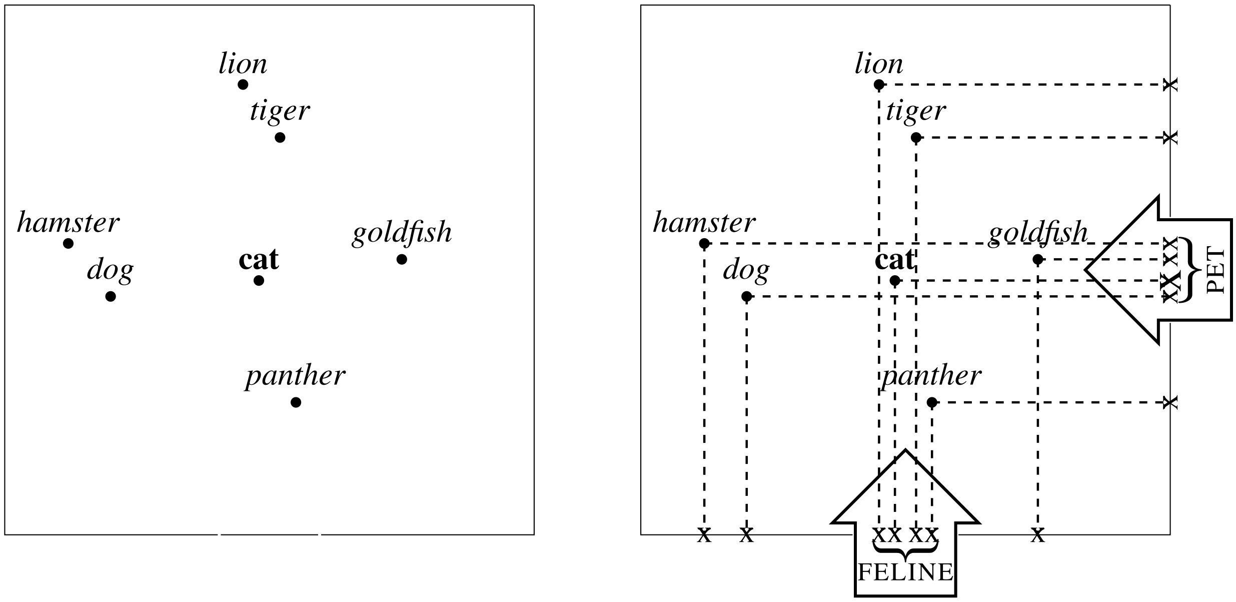
4 Semantics in Perspective

An advantage of distributional semantic techniques is that they provide robust representations that can in principle be composed in open-ended ways to generate previously unobserved but nonetheless semantically situated combinations of words, aligning with the combinatorial propensity of natural language itself. In a space of word-vectors built based on an analysis of the way words co-occur in a corpus, every word shares values across the same underlying set of features. So, for instance, if we are modelling the words *cat* and *dog* based on observations of thoses words in sentences in a corpus, we might reasonably expect categorical hypernyms such as *pet* and *animal* to have a high co-occurrence for both words, as well as descriptors such as *big*, *black*, and *pet*. Over the course of many such observations, quantified into a vector the features of which correspond to statistics about these observations, we will gradually see the vectors $\overrightarrow{cat}$ and $\overrightarrow{dog}$ occupy a similar region of the semantic space we are constructing, and from this we may infer that *cat* and *dog* denote similar things.

A related disadvantage is that the construction of semantic representations incorporates information about the way in which words co-occur with one another without regard for the conceptual nuances inherent in the syntagmatic relationships of words in a sentence. This means that such a model cannot in any real sense be construed as interpreting the language which it encounters or any underlying conceptual commitments; it is simply building up representations of linguistic tendencies, and the proximity of semantically related lexical representations is an artefact of the associated labels having correspondingly similar tendencies in their usage, not of any underlying extrapolation of conceptual structure from an experience of the world. In other words, a computational application based on a distributional semantic methodology can not in any sense be said to actually read or understand a text.

Furthermore, as implicatures and word senses are naturally not acknowledged by the process of constructing distributional semantic representations, all the looseness and ambiguity of language in use is imported into the representational framework. In the case of metaphor in particular, we must assume that metaphor will be ubiquitous in any non-trivial collection of language, and so metaphoric senses of words will be built into the model to a degree roughly in correlation with their typicality. So, returning to our example involving cats and dogs, *cat* will reasonably be expected to co-occur with, in addition to words indicating animals, words like *cool* and *hep* associated with the metaphorical connotation of the word as someone connected to jazz music, while *dog* will, in the sense that the word can metaphorically refer to a villain, be observed in the vicinity of other words like *treacherous* and *cursed*. An illustration of the inherent messiness of a distributional semantic model is offered in Figure 1a, where, without even transgressing literal senses of the word, we can see that the space surrounding *cat*, in this two-dimensional projection, includes terms related to both the colloquial sense of membership of the category PET and the biological sense associated with the category FELINE.

Our methodology proposes to turn this apparent disadvantage into a productive feature of a metaphor generating system. The insight underlying our approach is that, in a high dimensional space,



(a) A Semantic Space

(b) Conceptual Perspectives

Figure 1: The image on the left shows a semantic space as learned from the analysis of word distributions across a corpus, with all its inherent messiness, while the image on the right demonstrates how context specific conceptualisations are afforded by lower dimensional perspectives.

a variety of lower dimensional perspectives can be taken, affording the online projection of context specific semantic subspaces in which the effects of *ad hoc* conceptualisation take on momentary geometric characteristics. This idea is illustrated in Figure 1b, where we see that the context specific sense of the representations associated with *cat* collapse into two different coherent regions from two distinct points of view. This regionalisation of concepts is broadly in line with the *conceptual spaces* of Gärdenfors (2000), who describes a model by which conceptual domains are represented as convex regions in spaces delineated by dimensions that map to the intensional properties. The process of contextualisation in our framework and the conceptual geometries that emerge from these contextualisations are particularly relevant to the modelling of metaphor, since metaphor arises in context and can involve the projection of salient properties that are likewise contextually specified from one domain to another.

The procedure for building our context sensitive distributional semantic methodology begins with the selection of a corpus which will serve as a basis for analysising the ways in which words co-occur with one another. We've chosen to use the English language version of Wikipedia, in part because it contains a large number of well formed sentences covering a wide ranging vocabulary. We build a matrix of co-occurrence statistics tallying the number of times words occur within the same sentence and within a certain perimeter of one another, weighted using an information theoretical metric that takes into account the relative independent frequencies of words in order to avoid inflating the values associated with very common words. (technical details can be found in McGregor et al., 2015). The precise formulation of this weighting scheme, proposed in McGregor et al. (2015), is as follows:

$$M(w,c) = \log_2 \left(\frac{f(w,c) \times W}{f(w) \times (f(c) + a)} + 1 \right) \tag{1}$$

Here, $M(w,c)$ represents the weighted statistical value of a word a co-occurreing with a word c; $f(w,c)$ is the frequency at which w and c are observed to co-occur; $f(w)$ and $f(c)$ are the independent frequencies at which w and c occur in the corpus; and W is the overall count of word tokens in the corpus. The value a is a weighting constant designed to avoid the selection of especially obscure words in our context projection procedure (see Section 5), while 1 is added to the overall ratio in order to ensure that all values are greater than zero, with a value of zero indicating that w and c have never been observed to co-occur at all.

This initial modelling procedure results in a large and sparse array of statistics about language use—large in that there are about eight million unique word types in Wikipedia, any one of which might be observed co-occurring with any other word, and sparse in that most of these never do co-occur with one another, and indeed a great many of these words only appear once or twice across all five-million-plus articles on the website. The values of this large, sparse matrix are effectively coordinates for situating semantic representations in a space. With this in mind, we treat this matrix as a *base space* from which we can perform contextualised semantic projections: this base corresponds to the messy, ambiguous semantic space of Figure 1a. An important difference between that visualisation and the computational operation of our methodology, however, is that there are a very large number of perspectives to be taken on a high-dimensional vector space. So, for instance, if we were to consider all the possible ways to pick a subspace from an eight million dimensional base space, we would have to sort through a total of $2^{8,000,000}$ projections—a figure significantly larger than the number of atoms in the entire universe.

This exponential explosion of possible subspaces has sometimes been referred to as *the curse of dimensionality* by statisticians and computer scientists, as it leads to problematically immense search spaces when, for instance, trying to find an optimal combination of parameters for solving a problem (Bellman, 2003). In the case of our methodology, however, we see this proliferation of potential geometric arrangements between word-vectors as a source of contextual bountifulness, providing the basis for an effectively limitless range of conceptual relationships. Having established a base space of lexical representations that permit the establishment of a preponderance of semantic perspectives, the crucial question becomes how we choose the right subspace for discovering and thereby generating a new metaphor.

5 Projecting Metaphorical Mappings

In principle, our contextual projections should correspond with a specific conceptualisation of the world, and so also with some cognitive state: the projection represents the situation in which a cognitive-linguistic agent finds occasion to adapt a word to a metaphorical context. We would like to be able to somehow preemptively select a range of co-occurrence dimensions which capture this process of conceptualisation and then feel our way through this subspace, discovering the mappings between semantic representations as opportunities for metaphoric meaning-making. By this view, metaphor generation is a process of production and then search, entailing the perception on the part of the metaphor-making agent of ways to mean and to communicate meaning in a cognitive environment. The methodology is therefore in theory a dynamic one, with input from the world serving as the basis for a projection which in turn lends itself to the output which completes the loop between agent and environment.

In practice, for the purposes of this exposition of a novel computational framework, we will perform something of a trick of reverse engineering, beginning with metaphors each mapping a range of semantic representations from one conceptual domain to another and then examining the subspaces that facilitate these mappings. This process will be, in effect, a proof-of-concept: we will demonstrate that there are, in principle subspaces in which metaphors are resolved as mappings between geometrically aligned regions, and that these subspaces can be interpreted in terms of a set of observable co-occurrence features across which the components of metaphors have regular tendencies over the course of our corpus. What will remain an open question is whether there is a systematic way to select these co-occurrence features based on a particular contextual input, without a metaphor mapping from words in a source domain to words in a target domain specified *a priori*.

In order to construct these subspaces, then, we begin with a group of words comprised of a set of metaphorical sources paired with a set of metaphorical targets. We examine the word-vectors for each of the words in this set and extrapolate a range of co-occurrence dimensions with relatively high values for all the words in the set. We can consider these dimensions to broadly delineate a semantic subspace capturing the features salient to the conceptual category implied by that particular combination of input words. So, for instance, taking the target words (*surgeon, scalpel, patient, hospital*) and the source words (*butcher, knife, meat, shop*) as our inputs, we discover that the collectively most typical co-occurrence terms for these words are things like *nurse*, *stab*, *cutting*, and *kill*.

Next, having established this set of co-occurrence dimensions which we know to be informative about all the input terms, we can iterate through combinations of these dimensions in order to discover the sets that are best suited to projecting the pairs of metaphors as isomorphically aligned congruent geometric regions. Four examples of such projections, drawing from a set of 100 dimensions deemed conceptually relevant to the word-vectors being analysed and using combinations of only three co-occurrence dimensions for the sake of visualisation, are presented in Figure 2.[2] In each plot, two different conceptual domains are represented as two separate quadrilaterals, each corner being defined by the position of a word-vector labeled with a word that is semantically pertinent to that domain. In each case, by extending a line from the origin of the subspace through and then beyond a point in the source domain, we very nearly arrive at the point indicated by the corresponding representation in the metaphorical target domain. In each case, the subspace has been chosen in order to facilitate this generative mapping.

Plots 2a and 2b map from relatively concrete domains, pertaining to vision and proprioception respectively, to the more abstract domains of emotion and quality judgments. As Lakoff and John-

[2]Note that the curse of dimensionality is still at play here, if not quite at the same scale as when picking arbitrary subspaces from the full base space of 8 million co-occurrence features: there are 161,700 ways to pick three dimensions from 100.

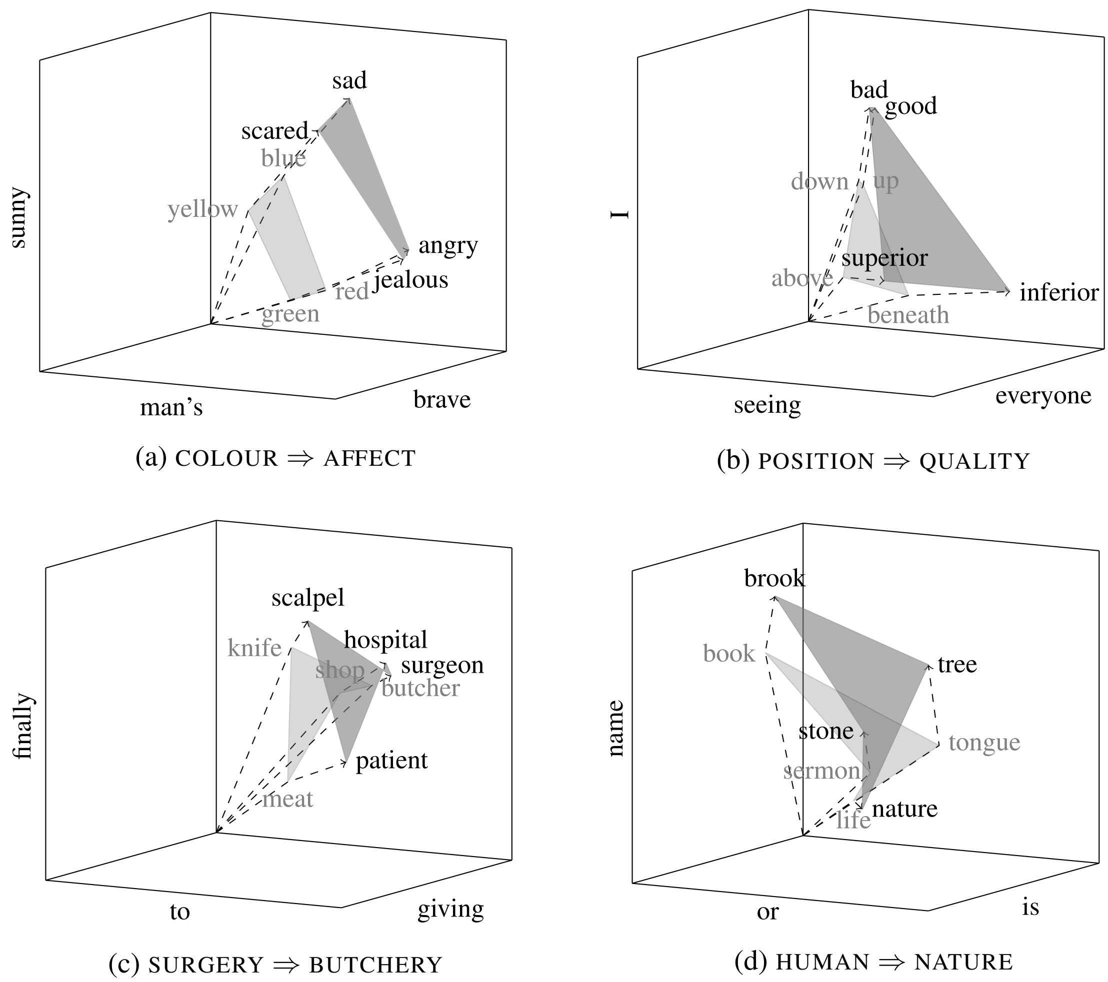


Figure 2: Four examples of context specific projections in which metaphoric mappings between conceptual regions of semantic representations emerge.

son (2003) have written, metaphor "involves all the natural dimensions of our experience, including aspects of our sense experiences," (p. 235), and so these colour and orientation metaphors fit neatly into the general ontology of the theory of conceptual metaphors. They are, by the same token, tinged with cultural relativism: as Levinson (1996) tells us, the cognitive construction and corresponding vocabulary of spatial relations in particular can vary immensely between different groups of language users.[3] With this in mind, it should be possible to discover subspaces in which these mappings would work out differently. The point to appre-

[3] An interesting question to pursue empirically would be to see if particular corpora, which, as Caliskan et al. (2017) have recently pointed out, are themselves prone to absorb culturally specific biases that are then subsequently necessarily incorporated into statistically derived semantic models, are more or less conducive to the modelling of corresponding cultural metaphors.

ciate here, though, is that these different subspaces correspond to different conceptual contextualisations, and in principle we should be able to generate any metaphor we like in an open-ended way: after all, in some conceptual sense, ravens can be writing desks and people are telescopes, and we would like our methodology to afford us the opportunity to discover all of these mappings.

Plots 2c and 2d, on the other hand, are mappings between more specific domains. In the first instance, we examine a metaphor that has been discussed perhaps to the point of tedium by linguists, namely that of the butcherous surgeon, filled in with some more details regarding potential intensional mappings from the one profession to the other. Something to note here is the potential for reversal of the metaphor, since we can easily enough imagine describing a particularly skillful butcher with the vocabulary of stereotypical surgical precision. But again, as with the metaphors of orientation and colour, a recalibration of these relationships would involve a reconsideration of the underlying dimensions defining the subspace, and so in effect a reconceptualisation of the cognitive context in which the new metaphor would arise. This potential for reversibility gives rise to another interesting question, however, since not all metaphors are evenly balanced in this regard: to borrow an example from Veale and Hao (2007), the metaphor *cigarettes* are *time bombs* is easy enough to understand, but swapping source and target generates something more conceptually bizarre. The point is not that there should be no way to reverse any given metaphor, because there is certainly *some* context in which *time bombs* are *cigarettes*, but this mapping requires a greater degree of contextualisation—and so we expect the appropriate subspace to be less likely to arise.

Finally, in the case of Plot 2d, we have generated an interpretation of Duke Senior's speech at the beginning of the Second Act of *As You Like It*, where he declares:

> And this our life, exempt from public haunt,
> Finds tongues in trees, books in the running brooks,
> Sermons in stones, and good in everything.

Here we have an example of a subspace in which we can generate the evocative kind of mappings that Carston (2010) has referred to as *imagistic* in their propensity for conjuring up non-propositional mental content. But, where Carston proposes two distinct, albeit not exclusive, routes to metaphor understanding, we would like to present a methodology which encapsulates a singular path to metaphor generation. So here we see that, in the contextual projection of geometric subspaces, we can frame metaphor generation in terms of a unified process of mapping between isomorphic domains: the metaphor is, in effect, the product of the perception of *how* to convey mental content that has arisen in response to some situation in the world, and the discovery of the words that do the job under those circumstances is not really any different whether the metaphor is conveying something banal or sublime.

In every case explored here, we suggest that the conceptually specific subspace of contextualised semantic relationships offers up affordances of meaning making, and the language user grabs these lexical opportunities for communication and applies them, much as if they were a tool found lying on the ground at just the right moment. Our case is that, in addition to accommodating the *ad hoc* construction of conceptual structures, a geometric approach involving congruences within contextually projected subspaces can create a basis for modelling the evident perceptual immediacy with which metaphor is generated. Once a subspace is specified, identifying mappings between domains is simply a matter of identifying aligned and matched shapes. The outstanding question is then how to determine these subspaces in the first place. In the following section, we will explore a technique for discovering these subspaces analytically, and will suggest the broad possibility for a data driven approach for the generative construction of such subspaces.

6 Finding Coherent Subspaces

In the subspaces illustrated in Figure 2, the idea is that metaphor correlates with a mapping by which the angle between the source vector and the target vector is small, and the overall variances in distances between source vectors and respective target vectors is likewise small. This results in a convenient way of both visualising and conceptualising these cross-domain mappings as projections within projections, the constellation of word-vectors outlining the target domain being like a shadow of the configuration of points in the source domain cast by a light source located at the origin of the subspace. Since these subspaces are composed of statistics construed along concrete dimensions of co-occurrence as observed in a corpus, we can interpret this phenomenon mathematically in terms of sets of dimensions with comparable relative probabilities of co-occurrence for source versus target terms. This means the ratio of co-occurrences for a source and target word with a particular term should be relatively consistent with the ratio of sums of co-occurrences across all terms delineating the dimensions of a contextualised mapping. In mathematical terms, we are trying to find the set of d co-occurrence features c that minimise the variance in the ratios of the co-occurrence measure M from Equation 1 for a given source word w_s and target word w_t:

$$\sum_{j=1}^{d}\left(\frac{M(w_s,c_j)}{M(w_t,c_j)}-\mu\left(\frac{M(w_s,c)}{M(w_t,c)}\right)\right)^2=0 \qquad (2)$$

Here, c_j indicates any one of the d co-occurrence features in the subspace projected to represent the contextual emergence of a metaphor, and μ indicates taking the mean of the ratios of the measures M for source word to target word for all such features.

This formalism will be useful when we consider how we might go about solving the problem of systematically specifying the context of a metaphor in terms of co-occurrence statistics. More generally, the geometry of an appropriately configured subspace captures the semantics inherent in relationships between lexical representations. In the case of a mapping from a source to a target in one of these spaces, we can say that there is a specific direction, distance, and orientation associated with the metaphoric move from one conceptual domain to another, and these features in turn correlates with a probabilistic tendency in terms of the way that language has been observed in use.

One important thing to note about all the plots in Figure 2 is that the co-occurrence dimensions that form the basis for these mappings have an evidently random character. One would not instinctively guess, for instance, that co-occurrences with the words (*sunny, man's, brave*) had something to do with a structured mapping from the domain of COLOUR to that of AFFECT. What we must bear in mind, though, is first of all that these dimensions collectively represent a tendency in language use, and so should not necessarily be interpreted independently. Instead of thinking of the dimensions defining a concept-specific context as a set of labels denoting relevant conceptual properties, we propose that these dimensional tags should be taken as in some sense collectively topical, and potentially in ways that are apparent in the overall dynamic of dimensions rather than in the aggregate of individual dimensions. In this regard, we present our framework in alignment with the pragmatic perspective on metaphor as a lexical specification in response to a context in an environment rather than as a structured mapping between persistent conceptual representations.

The second thing to appreciate with regard to the definition of

source	*target*	*cosine*	*dimensions*	*source*	*target*	*cosine*	*dimensions*
blue	sad	0.96	*beautiful, sees,*	down	bad	0.97	*toughness, to,*
yellow	scared	0.94	*girl, confused,*	up	good	0.96	*it's, feeling,*
red	angry	0.90	*staring, looks,*	above	superior	0.95	*rating, bone,*
green	jealous	0.92	*upset, pretty*	beneath	inferior	0.95	*vein, nerves*
knife	scalpel	0.91	*performed, arm,*	book	brook	0.89	*referring, is,*
shop	hospital	0.96	*nose, removed,*	tongue	tree	0.92	*bear, beneath,*
butcher	surgeon	0.90	*stanley, safety,*	sermon	stone	0.93	*creates, trees,*
meat	patient	0.89	*informs, warren*	life	nature	0.98	*green, itself*

Table 1: Mappings, angles, and labels of axes for eight dimensional subspaces of the same metaphor generating projections described in Figure 2.

a contextualised subspace is that the examples illustrated in Figure 2 have been projected into just three dimensions, and the decision to use such a low number of dimensions was taken purely for the sake of visualising the way that metaphoric relationships play out in terms of geometric mappings. We can at least in principle, though, select arbitrarily large numbers of dimensions for our projections, and a reasonable hypothesis would be that, as we move into higher dimensional subspaces, the mappings between our conceptual regions of word-vectors become more discriminating, because we can select a profile of dimensions that weeds out more of the noise which creeps into mappings in lower dimensional projections. As we increase the dimensionality of a subspace, we might expect the set of co-occurrence features defining the subspace to take on a conceptually coherent characteristic: with the ratios of lengths of source and target word-vectors spread out across a larger number of dimensions, there is more room for discovering co-occurrences that are specific to components of the context in which the metaphor is generated.

This expectation bears out to at least some extent in Table 1, where the same four metaphors explored in Figure 2 are reconstructed in eight dimensional subspaces. Here *cosine* measures the angle between source and target, with scores of zero indicating the two vectors are at right angles while scores of one point to perfect alignment. The consistently high cosine scores suggest that these higher dimensional spaces are still doing the generative work of mapping source to target, while the dimension labels associated with each projection now take on an at least qualitatively more coherent aspect. The dimensions associated with the COLOUR to AFFECT metaphor, for instance, are indicative of vision, aesthetics, and emotions, while the dimensions for the SURGERY to BUTCHERY mapping are, taken collectively, about doing things to bodies.[4]

In these cases, the eight dimensions have been selected from all the possible combinations of just the top 25 dimensions in terms of mean co-occurrence statistics for the words involved in the metaphor: as the number of dimensions involved in a projection increases, the computational cost of computing all combinations of that cardinality from a base set of dimensions expands exponentially.[5] This means that there might be scope for discovering even more productive combinations of dimensions amongst a broader set of base co-occurrence dimensions. More to the point, though, it also raises the important question of whether there might be a way to systematically predict the subspaces in which particular metaphorical mappings will unfold. More than just overcoming the well-established problem of combinatory searches through high-dimensional parameter spaces, this would point the way towards a truly generative model, where simply the representational components combined with a conceptually anchored input that correctly contextualises these representations provides the basis for a projection and a corresponding mapping coupling target word-vectors with metaphorical source word-vectors.

One way forward would be to treat metaphor generation as a

[4]The presence of the presumably decapitalised proper names *stanley* and *warren* is intriguing, if not immediately explicable—though a basic search of the underlying corpus, which is to say, Wikipedia, does indeed indicate various surgeons known by either name, and various other axes of association between these names and numerous senses of the word *butcher*.

[5]There are 1,081,575 ways to choose 8 dimensions from a set of 25 dimensions: the curse of dimensionality is still at play here, if not at quite the same scale as when projecting arbitrary subspaces from a base space of 8 million dimensions.

machine learning task. So, for instance, given a target such as *surgeon*, a context such as *reckless*, and a set of attributes associated with the target such as *scalpel* and *patient*, we could seek to train a model to learn to pick the co-occurrence features defining a subspace in which the vectors within the domain of SURGERY map to the corresponding points within the domain BUTCHERY. This would constitute a *supervised* machine learning problem, in that we would train the model based on a set of known cross-domain mappings, coupled with keys indicating the context in which these particular mappings are activated. This research would therefore entail the development of a fairly large dataset of such annotated mappings, which is not a trivial undertaking, though work on cataloging metaphor described by, for instance, Steen et al. (2010) and Stickles et al. (2016) offers a good starting point. The model itself could take the form of a neural network, and the criterion for defining a good mapping outlined in Equation 2 presents a candidate objective function.

For now, returning to the theoretical objective of exploring a computational implementation of a pragmatic account of metaphor outline in Section 2, we suggest that the prospective methodology we have outline here could be construed as capturing the way that *ad hoc* concepts are formed in the process of metaphor generation, in response to a particular context in an agent's environment. The particular arrangement of the word-vectors associated with, for instance, BUTCHERY or EMOTION in the source domains illustrated in Figure 2 indicate a context specific projection of those concepts, and one which facilitates a productive cross-domain correspondence. The curse of dimensionality remains a factor here, in that there are a very large number of ways to rearrange the semantic geometry of a collection of lexical representations, but this has to be the case: we must allow our methodology for discovering metaphors to be open-ended in the way that different situations specify different extemporaneous conceptual constructs. We have so far demonstrated that there are some steps that can be taken to reduce the immensity of the space of possible semantic projections based on a set of inputs. Whether we can systematically identify a methodology for representing context and then identifying the particular subspaces in which a specific context is resolved into a set of geometric relationships that can be acted upon as the basis for linguistic communication remains to be seen.

7 The Way Forward

It is important at this point to consider ways in which a methodology such as the one we've proposed might be quantitatively evaluated. One approach to evaluating metaphor generation might involve a statistical anlaysis of the correlation between the words involved in a novel construction and the anticipated affect of those words, as Veale and Hao (2007) have done for similes. This is an interesting and relevant way of examining the impact of figurative language, but it doesn't really tell us about the way that this type of language operates as a mechanism of communication. Alternatively, Miyazawa and Miyao (2017) have proposed to crowd-source human assessments of computationally generated metaphors in terms of metrics such as *novelty* and *comprehensibility*. This grounds a system out in a real-world judgement of the degree to which a metaphor accomplishes the task of communicating a proposition in a new way, but it tells us nothing else about how metaphor operates in a particular communicative context.

If we are to take seriously Davidson's (1978) idea that metaphor is not strictly to be assessed in terms of the success of a non-literal construction in conveying propositional content, we can't even really query a pragamatically motivated system's output in terms of entailment: when an agent constructs a metaphor, they are doing something aside from, or more than, simply saying that something is a certain way in the world. From this point of view, it seems it will be necessary to develop an approach to measuring the consequence of an expression in a particular circumstance, beyond simply determining whether propositional information has been effectively communicated. It seems inevitable that, in order to evaluate the context-sensitive generation of metaphor, we will need to devise a way to place the evaluator directly into the particular context in which the metaphor has come about. What this would involve, and the degree of immersion necessary to really evaluate the system, depends on how exactly context is defined, a question which is fundamental to everything that has been said here. The interpretability of the mechanisms of the methodology outlined in the preceding sections might offer some leverage in terms of providing insight into the operation of the system, something which is a factor for artificial agents in a way it is not for humans.

It is also worth considering how we might establish a baseline as a point of comparison to any system designed based on the methodology outlined above, and here we should note that a new generation of context-sensitive distributional semantic models have emerged which seem ripe for application to non-literal language. Using, for instance, bi-LSTM (Peters et al., 2018) or transformer architectures (Devlin et al., 2019), massively deep and wide neural networks learn to project word representations into a semantic space that is coloured by the context of an actual sentence in which the word is observed. On the surface, models such as this would seem to do something very similar to what we are grasping for with our proposed methodology. But, while some valuable work has been done to explore the geometries of these types of models and the ways in which representations shift in context (Ethayarajh, 2019), the question of how these approaches might push words into geometries that can be interpreted in terms of context sensitive departures from literal lexical interpretations remains open.

So, in terms of next steps, we would propose first of all considering a robust approach to evaluating the way that metaphor comes about in a particular environmental context, and then exploring ways in which both output and process can be examined in order to assess the operation of a metaphor in a situation. These are substantial research objectives in themselves, and in addition to the significant work that remains to be done in refining the broad methodology proposed here.

8 Conclusion

For the time being, we will have to consider the prospect of a comprehensive computational metaphor generating system described above as a target for future work. The methodology for mapping sources to targets outlined in this paper has, we hope, put some of the fundamental apparatus for such a system in place, but there are still a number of questions regarding the technical particulars of the approach. Model parameters range from the way in which words are represented (should they be lemmatised, for instance, or endowed with syntactic information?) to the way that frequentist statistics are weighted in our base matrix (is the information theoretic measure used to obtain the case studies outlined here the most appropriate one?). The establishment of the features of the relationships between the vectors associated with input words and the candidate dimensions for contextually projecting these vetors bears a great deal more thought, and indeed the selection of the underlying data itself is a question requiring careful theoretical consideration.

As things stand, we've endeavoured to present a procedure that demonstrates the tractability of spaces of statistically grounded semantic representations as geometric models of the process of metaphor generation. To anticipate one potential and reasonable theoretical objection to the framework we've outlined here, much of the work seems to be happening on the surface of language, by way of the interaction of representational forms without necessarily communicating with associated cognitive processes that bind these forms to non-linguistic modes of being in the world. Ultimately, important reservations have been raised by, for instance, Chomsky (1957) about the very idea that much of what is inherent in language can be gleaned from the statistical data about how words have been used, and we acknowledge that there is much to word meaning which is simply not in the words themselves. This is a valid point, and we would be remiss to argue that the intentional and indeed phenomenological properties that have, at times, been associated with metaphor use could in any comprehensive sense be captured by a model based on our methodology. Ultimately, it is at least admissible to surmise that the contextualisation component of our methodology in particular will have to fall back on information that comes in some sense from outside the statistical data about language use. There is, to put it concretely, something in the lived-in experience of the world that differentiates bad surgeons from, say, bad barbers, and this distinction has significant implications for the way that metaphor might be applied.

Nonetheless, we maintain that there is something desirable in the application of statistical geometry in performing mappings between semantic domains. The first desideratum has been a core theme of the description of our methodology above: by seeking semantics in a geometry of quantitative lexical representations, we likewise effectively map the operation of our methodology to underlying claims about the way that language is observed in use, and so pin it at least in a general sense to the primary criterion we have for talking about what language is used for and how it works. We have remained intentionally vague about what these claims might actually be, but we would like to create a framework for exploring the pragmatic commitment that language is in some sense in the world as much as it is in the head. A requirement of such a framework would be a mechanism for the *ad hoc* generation of concepts, in a process where language is both an expressive product and also a productive source of the situational specification of meaning. This process can be particularly evident in the case of metaphor generation, where extemporaneously constructed concepts can vary in surprising ways from the prototypical or stereotypical versions of those concepts.

As we have mentioned, the project of devising a computational framework for lexical pragmatics is an ambitious one: this area has typically been viewed as recalcitrant to computational modelling, not least because it deals with the inherent messiness of being an agent in a chaotic world, and there being something it is like to be such an agent. We propose that, by cautiously considering the rampant recombinability of high dimensional spaces, we might discover a basis for using statistical analysis as at least a simulacrum of the complexity of the world. In the cases illustrated above, based on the weightings used to build our base matrix and the nature of the mappings from source to target domains, the methodology is picking up on something about consistencies in relative probabilities between coupled representations as well as across sets of these couplings, but there is no reason we can't consider any range of other hypotheses about what tendencies in language use, or indeed complex interrelationships between tendencies, might be associated with metaphor. We will simply note that it is probably desirable to consider models where the features of representational structures are in some sense interpretable: this will facilitate training a system to make the selections necessary in order to project contextualised versions of these representations.

The second desideratum regards the construction of a framework that permits the notion of perceptual affordance to play a role in the modelling of linguistic flexibility, something we touched on at the end of Section 5 when we discussed the perceptual immediacy of modelling metaphors geometric features. By designing a methodology in which conceptual regions can be mapped to each other through congruences in their directly observable geometric situations, we can begin to see a way towards generating metaphor without necessarily having to do the more cognitively expensive work of simultaneously interpreting metaphor. This is visually obvious in the examples illustrated in Figure 2, where we can simply see that the source domains project themselves neatly onto the target domains, and this observation extends intuitively to higher dimensions, where this visual immediacy can be generalised in terms of some very simple formulations regarding angles and distances. Our claim is that this methodology for projecting spaces in which metaphors play out geometrically instantiates a theoretical commitment that metaphor can be described as the direct perception of an opportunity for meaning-making in a specific cognitive environment.

Under this regimen, metaphor becomes not just a mechanism for codifying propositional content about the world in ways that are by turn aesthetically appealing and communicatively efficient, but also an instance of humans using words in the same way that they use other entities, however material or immaterial they may be, found in their environments. Just as a shoe can become a hammer or a newspaper can become a fan without the agent deploying these implements in those ways thinking too much about it, a word can become a device for showing not just what but how a communicator is cognitively experiencing their situation. This does not dispense with the notion of metaphor as a property of the productive and often systematic cognitive coupling of conceptual domains; on the contrary, unless we choose to likewise dispense with the idea that language is tightly coupled with cognition, this becomes another piece of evidence in the ever more clear picture that words offer great insight into the working of the mind. In fact, we would like to believe that our methodology, to the extent that it offers a way of seeing how concepts that come about contextually can still be the subject to cross-domain mappings, and how furthermore statistics about language use can become the basis for the geometry in which these mappings unfold, can be an important component in the application of computational techniques to the study of the relationship between language, humans, and the world.

References

Agres, K. R., S. McGregor, K. Rataj, M. Purver, and G. A. Wiggins (2016). Modeling metaphor perception with distributional semantics vector space models. In *Workshop on Computational Creativity, Concept Invention, and General Intelligence*.

Barnden, J. A. and M. G. Lee (1999). An implemented context system that combines belief reasoning, metaphor-based reasoning and uncertainty handling. In *Modeling and Using Context: Second International and Interdisciplinary Conference*, pp. 28–41.

Baroni, M. and R. 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.

Barsalou, L. W. (1992). Frames, concepts, and conceptual fields.

In A. Lehrer and E. F. Kittay (Eds.), *Frames, Fields, and Contrasts: New Essays in Semantic and Lexical Organization*, pp. 21–74. Hillsdale, N.J.: Lawrence Erlbaum Associates.

Barsalou, L. W. (1993). Flexibility, structure, and linguistic vagary in concepts: manifestations of a compositional system of perceptual symbols. In A. Collins, S. Gathercole, and M. Conway (Eds.), *Theories of memory*, pp. 29–101. London: Lawrence Erlbaum Associates.

Bellman, R. E. (2003). *Dynamic Programming*. Dover Publications, Incorporated.

Caliskan, A., J. J. Bryson, and A. Narayanan (2017). Semantics derived automatically from language corpora contain human-like biases. *Science 356*(6334), 183–186.

Carston, R. (2002). Metaphor, ad hoc concepts and word meaning - more questions than answers. *UCL Working Papers in Linguistics 14*, 83–105.

Carston, R. (2010). Metaphor: Ad hoc concepts, literal meaning and mental images. *Proceedings of the Aristotelian Society 110*(3), 297–323.

Carston, R. (2012). Metaphor and the literal/nonliteral distinction. In K. Allan and K. M. Jaszczolt (Eds.), *The Cambridge Handbook of Pragmatics*, pp. 469–492. Cambridge University Press.

Chomsky, N. (1957). *Syntactic Structures*. The Hague: Mouton and Co.

Clark, A. (2006). Language, embodiment, and the cognitive niche. *Trends in Cognitive Sciences 10*(8), 370–374.

Clark, S. (2015). Vector space models of lexical meaning. In S. Lappin and C. Fox (Eds.), *The Handbook of Contemporary Semantic Theory*, pp. 493–522. Wiley-Blackwell.

Coecke, B., M. Sadrzadeh, and S. Clark (2011). Mathematical foundations for a compositional distributed model of meaning. *Linguistic Analysis 36*(1-4), 345–384.

Davidson, D. (1978). What metaphors mean. In *Inquiries into Truth and Interpretation* (2nd ed.). Oxford: Clarendon Press.

Devlin, J., M.-W. Chang, K. Lee, and K. Toutanova (2019). BERT: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics*, pp. 4171–4186.

Ethayarajh, K. (2019). How contextual are contextualized word representations? comparing the geometry of BERT, ELMo, and GPT-2 embeddings. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing*, pp. 55–65.

Fauconnier, G. and M. B. Turner (2008). Rethinking metaphor. In R. Gibbs (Ed.), *Cambridge Handbook of Metaphor and Thought*. Cambridge University Press.

Gärdenfors, P. (2000). *Conceptual Space: The Geometry of Thought*. Cambridge, MA: The MIT Press.

Gargett, A. and J. Barnden (2013). Gen-Meta: Generating metaphors using a combination of AI reasoning and corpus-based modeling of formulaic expressions. In *Proceedings of TAAI 2013*.

Gentner, D. (1983). Structure-mapping: A theoretical framework for analogy. *Cognitive Science 7*(2), 155–170.

Gibbs Jr., R. W. (1994). *The Poetics of Mind*. Cambridge University Press.

Gibbs Jr., R. W. and M. Tendahl (2006). Cognitive effort and effects in metaphor comprehension: Relevance theory and psycholinguistics. *Mind and Language 21*(3), 379–403.

Gibson, J. J. (1979). *The Ecological Approach to Visual Perception*. Boston: Houghton Miffline.

Grefenstette, E., M. Sadrzadeh, S. Clark, B. Coecke, and S. Pulman (2014). *Concrete Sentence Spaces for Compositional Distributional Models of Meaning*, pp. 71–86. Dordrecht: Springer Netherlands.

Grice, H. P. (1975). Logic and conversation. In P. Cole and J. L. Morgan (Eds.), *Syntax and Semantics Volume 3: Speech Acts*, pp. 41–58. New York: Academic Press.

Gutiérrez, E. D., E. Shutova, T. Marghetis, and B. K. Bergen (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.

Kartsaklis, D. and M. Sadrzadeh (2013). Prior disambiguation of word tensors for constructing sentence vectors. In *Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing*, pp. 1590–1601.

Kintsch, W. (2000). Metaphor comprehension: A computational theory. *Psychonomic Bulletin & Review 7*(2), 257–266.

Lakoff, G. and M. Johnson (2003). *Metaphors We Live By* (2nd ed.). University of Chicago Press.

Levinson, S. C. (1996). Relativity in spatial conception and description. In J. J. Gumperz and S. C. Levinson (Eds.), *Rethinking linguistic relativity*, pp. 177–202. Cambridge University Press.

Mason, Z. J. (2004). Cormet: A computational, corpus-based conventional metaphor extraction system. *Computational Linguistics 30*(1), 23–44.

McGregor, S., K. Agres, M. Purver, and G. Wiggins (2015). From distributional semantics to conceptual spaces: A novel computational method for concept creation. *Journal of Artificial General Intelligence*.

McGregor, S., E. Jezek, M. Purver, and G. Wiggins (2017). A geometric method for detecting semantic coercion. In *Proceedings of 12th International Workshop on Computational Semantics*.

Mikolov, T., K. Chen, G. Corrado, and J. Dean (2013). Efficient estimation of word representations in vector space. In *Proceedings of ICLR Workshop*.

Miyazawa, A. and Y. Miyao (2017). Evaluation metrics for automatically generated metaphorical expressions. In *12th International Workshop on Computational Semantics*.

Pantel, P. (2005). Inducing ontological co-occurrence vectors. In *Proceedings of the 43rd Annual Meeting on Association for Computational Linguistics*, Stroudsburg, PA, USA, pp. 125–132. Association for Computational Linguistics.

Peters, M., M. Neumann, M. Iyyer, M. Gardner, C. Clark, K. Lee, and L. Zettlemoyer (2018). Deep contextualized word representations. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics*, pp. 2227–2237.

Reimer, M. (2001). Davidson on metaphor. *Midwest Studies in Philosophy 25*, 142–155.

Reimer, M. (2013). Grice on irony and metaphor: Discredited by the experimental evidence? *International Review of Pragmatics 5*(1), 1–33.

Salton, G., A. Wong, and C. S. Yang (1975). A vector space model for automatic indexing. In *Proceedings of the 12th ACM SIGIR Conference*, pp. 137–150.

Schütze, H. (1992). Dimensions of meaning. In *Proceedings of the 1992 ACM/IEEE conference on Supercomputing*, pp. 787–796.

Searle, J. R. (1979). Metaphor. In A. Ortony (Ed.), *Metaphor and Thought*. Cambridge University Press.

Shutova, E. (2015). Design and evaluation of metaphor processing systems. *Computational Linguistics 41*(4), 579–623.

Shutova, E., S. Teufel, and A. Korhonen (2013). Statistical metaphor processing. *Computational Linguistics 39*(2), 301–353.

Steen, G. J., A. G. Dorst, J. B. Herrmann, A. Kaal, T. Krennmayr, and T. Pasma (2010). *A Method for Linguistic Metaphor Identification: From MIP to MIPVU*. Converging Evidence in Language and Communication Research. Amsterdam: John Benjamins Publishing Company.

Stickles, E., O. David, E. K. Dodge, and J. Hong (2016). Formalizing contemporary conceptual metaphor theory: A structured repository for metaphor analysis. *MetaNet, Special Issue of Constructions and Frames 8*(2).

Sweetser, E. (1990). *From Etymology to Pragmatics: Metaphor and Cultural Aspects of Semantic Structure*. Cambridge University Press.

Tsvetkov, Y., L. Boytsov, A. Gershman, E. Nyberg, and C. Dyer (2014). Metaphor detection with cross-lingual model transfer.

In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics*, pp. 248–258. The Association for Computer Linguistics.

van Genabith, J. (1999). Metaphors and type theory. In *Proceedings of the AISB'99 Symposium on Metaphor, Artificial Intelligence, and Cognition*, pp. 108–112.

Veale, T. and Y. Hao (2007). Comprehending and generating apt metaphors: A web-driven, case-based approach to figurative language. *AAAI*, 1471–1476.

Wilks, Y. (1978). Making preferences more active. *Artificial Intelligence 11*(3), 197–223.