

Re-Representing Metaphor: Modelling metaphor perception using dynamically contextual distributional semantics

Stephen McGregor^{1, 2*}, Kat Agres³, Karolina Rataj⁴, Matthew Purver⁵, Geraint A. Wiggins^{6, 5}

¹École Normale Supérieure, France, ²UMR8094 Langues, textes, traitements informatiques, cognition (LATTICE), France, ³Institute of High Performance Computing (A*STAR), Singapore, ⁴Adam Mickiewicz University in Poznań, Poland, ⁵Queen Mary University of London, United Kingdom, ⁶Vrije University Brussel, Belgium

Submitted to Journal:
Frontiers in Psychology

Specialty Section:
Cognitive Science

Article type:
Original Research Article

Manuscript ID:
413117

Received on:
09 Jul 2018

Revised on:
18 Mar 2019

Frontiers website link:
www.frontiersin.org

Conflict of interest statement

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest

Author contribution statement

SM - Lead author, primary architect of computational model

KA - Contributed to overall article, particularly psycholinguistic research

KR - Contributed writing in psycholinguistic section of paper, also responsible for the dataset we used

MP - Contributed to overall article, particular sections describing distributional semantic methods

GW - Contributed to overall article, particularly regarding modelling commitments and results

Keywords

distributional semantics, metaphor, Conceptual models, Computational Creativity, machine learning

Abstract

Word count: 164

In this paper, we present a novel context-dependent approach to modelling word meaning, and apply it to the modelling of metaphor. In distributional semantic approaches, words are represented as points in a high dimensional space generated from co-occurrence statistics; the distances between points may then be used to quantifying semantic relationships. Contrary to other approaches which use static, global representations, our approach discovers contextualised representations by dynamically projecting low-dimensional subspaces; in these *ad hoc* spaces, words can be re-represented in an open-ended assortment of geometrical and conceptual configurations as appropriate for particular contexts. We hypothesise that this context-specific re-representation enables a more effective model of the semantics of metaphor than standard static approaches. We test this hypothesis on a dataset of English word dyads rated for degrees of metaphoricity, meaningfulness, and familiarity by human participants. We demonstrate that our model captures these ratings more effectively than a state-of-the-art static model, and does so via the amount of contextualising work inherent in the re-representational process.

Ethics statements

(Authors are required to state the ethical considerations of their study in the manuscript, including for cases where the study was exempt from ethical approval procedures)

Does the study presented in the manuscript involve human or animal subjects: No

Re-Representing Metaphor: **Modelling metaphor perception using** **dynamically contextual distributional** **semantics**

Stephen McGregor^{1,*}, **Kat Agres**², **Karolina Rataj**^{3,4}, **Matthew Purver**⁵, and
Geraint Wiggins^{6,5}

¹*LATTICE (CNRS & École normale supérieure / PSL & Université Sorbonne nouvelle Paris 3 / USPC), 1, rue Maurice Arnoux, 92120 Montrouge, France*

²*Social and Cognitive Computing Department, Institute of High Performance Computing, A*STAR, Singapore, Singapore*

³*Faculty of English, Adam Mickiewicz University, Poznań, Poland*

⁴*Department of Cognitive Psychology and Ergonomics, University of Twente, Enschede, The Netherlands*

⁵*Cognitive Science Research Group, School of Electronic Engineering and Computer Science, Queen Mary University of London, Mile End Road, London E1 4NS, UK*

⁶*AI Lab, Vrije Universiteit Brussel, Pleinlaan 9, B-1050 Brussels, Belgium*

Correspondence*:
Stephen McGregor
semcgregor@hotmail.com

2 ABSTRACT

3 In this paper, we present a novel context-dependent approach to modelling word meaning,
4 and apply it to the modelling of metaphor. In distributional semantic approaches, words are
5 represented as points in a high dimensional space generated from co-occurrence statistics; the
6 distances between points may then be used to quantifying semantic relationships. Contrary to
7 other approaches which use static, global representations, our approach discovers contextualised
8 representations by dynamically projecting low-dimensional subspaces; in these *ad hoc* spaces,
9 words can be re-represented in an open-ended assortment of geometrical and conceptual
10 configurations as appropriate for particular contexts. We hypothesise that this context-specific
11 re-representation enables a more effective model of the semantics of metaphor than standard
12 static approaches. We test this hypothesis on a dataset of English word dyads rated for degrees
13 of metaphoricity, meaningfulness, and familiarity by human participants. We demonstrate that our
14 model captures these ratings more effectively than a state-of-the-art static model, and does so
15 via the amount of contextualising work inherent in the re-representational process.

16 **Keywords:** distributional semantics, metaphor, conceptual models, computational creativity

1 INTRODUCTION

17 Metaphor is a mode of re-representation: words take on new semantic roles in a particular communicative
18 context, and this phenomenon reflects the way that conceptualisation itself emerges during a cognitive
19 agent’s interaction with some situation in a dynamic environment. To describe someone as a *fox* will evoke
20 very different properties in a context which emphasises *cunning* and in one which emphasises *good looks*.
21 Metaphor, and the attendant transfer of intensional properties from one conceptual domain to another,
22 is therefore not just a matter of semantic encoding; rather, it involves an agent actually perceiving and
23 experiencing the world through a shift in conceptualisation, and correspondingly in cognitive and linguistic
24 representation.

25 Because metaphor occurs contextually, we hypothesise that the appropriate mode of lexical-semantic
26 representation will have some mechanism for contextual manipulation. With this in mind, we introduce
27 a methodology for constructing *dynamically contextual distributional semantic models*, allowing for the
28 *ad hoc* projection of representations based on the analysis of contextualising input. This methodology is
29 based on corpus-driven techniques for building lexical semantic representations, and the components of
30 these representations refer to observations about the way that words tend to occur with other words. The
31 ability to analyse these co-occurrence statistics dynamically will give our model the ability to generate
32 representations in the course of a developing, and potentially changing, conceptual context.

33 While the term *context* is often used in the field of natural language processing to refer explicitly to
34 the textual context in which a word is observed over the course of a corpus, our methodology has been
35 designed to capture something more in line with the sense of context explored by, for instance, Barsalou
36 (1999), who describes the way that a situation in an environment frames the context specific application of
37 a perceptually grounded symbol. Similarly, Carston (2010a) investigates the way that metaphor arises in
38 the course of the production of *ad hoc* concepts in reaction to a particular situation in the world. One of the
39 primary objectives of our methodology is to describe a framework that accommodates a pragmatic stance
40 on conceptual re-representation that is an essential aspect of metaphor.

41 In practice, we define contexts in terms of *subspaces* of co-occurrence features selected for their salience
42 in relation to a combination of input words. In the experiments described in the following sections, we
43 will seek to classify and rate the metaphoricity of verb-object compositions, using a statistical analysis
44 of the way that each word in the compositional dyad is observed to co-occur with other words over the
45 course of a large-scale textual corpus. So, for instance, if we have a phrase such as “cut pollution”, we
46 will build context-specific representations based on overlaps and disjunctions independently observed
47 in the co-occurrence tendencies of *cut* and *pollution*. These representations are *dynamic* in that they are
48 generated specifically in response to a particular input, and we show how this dynamism can capture the
49 re-representational quality by which metaphor is involved in the production of *ad hoc* concepts.

50 Importantly, our contextualisation methodology is not contingent on discovering actual collocations of
51 the words in a phrase, and in fact it is perfectly conceivable that we should be able to offer a quantitative
52 assessment of the metaphoricity of a particular phrase based on an analysis of a corpus in which the
53 constituent words never actually co-occur in any given sentence. This is because the representation of
54 a word dynamically generated in the context of a composition with another word is contingent on co-
55 occurrence features which are potentially shared between the words being modelled: while the words
56 *cut* and *pollution* could conceivably never have been observed to co-occur in a particular corpus, it is
57 very likely that they will have some other co-occurrences in common, and our methodology uses these
58 secondary alignments to explore contextual re-representations. We predict that it is not only the features of

59 the contextualised word representations themselves, but also the overall features of the subspace into which
60 they are projected (representing a particular conceptual and semantic context), which will be indicative of
61 metaphoricity.

62 A key element in the development of our methodology for projecting contextualised distributional
63 semantic subspaces is the definition of conceptual salience in terms of an analysis of specific co-occurrence
64 features. These features become the constituents of a geometric mode of metaphoric re-representation,
65 and our hypothesis is that a thorough analysis of the geometry of a contextually projected subspace will
66 facilitate the assessment of metaphoricity in context. The capacity for our model to make on-line selections,
67 as well as its susceptibility to replete geometric analysis, are key strengths that differentiate this from
68 existing quantitative techniques for representing metaphor. Our computational methodology is a variant of
69 an approach developed for context-dependent conceptual modelling (Agres et al., 2015; McGregor et al.,
70 2015); we describe the model and its application to modelling metaphor perception in Section 3.

71 The data that we use here to explore the re-representational capacities of our methodology consists
72 of human ratings of a set of English language verb-object phrases, categorised in equal parts as *literal*
73 non-metaphors, *conventional* metaphors, and *novel* metaphors, with each phrase given a rating by a group of
74 competent English speakers on a one-to-seven Likert scale for *metaphoricity* as well as for *meaningfulness*
75 and *familiarity*. We note that, in the context of this data (described in Section 4), metaphoricity has a
76 negative correlation with assessments of both meaningfulness and familiarity. In Section 5, we use this
77 data to train a series of regressions geared to learn to predict ratings for different semantic categories based
78 on the statistical geometry of subspaces contextualised by the concept conveyed by a given phrase.

79 Our methodology lends itself to a thorough analysis of the way different geometric features in a space
80 of weighted co-occurrence statistics indicate metaphoricity. One of our objectives is the extrapolation of
81 features that are particularly salient to shifts in meaning by way of conceptual re-representation, and to
82 this end we develop a methodology for identifying sets of geometric measures that are independently and
83 collectively associated with metaphor.

2 BACKGROUND

84 We have developed a novel computational model for metaphor processing, designed to treat metaphor as a
85 graded phenomenon unfolding in the context of an agent's interaction with a dynamic environment. In what
86 follows, we seek to ground our own model in research about the way humans process metaphor. This brief
87 survey leads on to a review of what have been some of the leading computational approaches to modelling
88 metaphor. Finally, we review the ways that existing computational approaches do and do not fit into our
89 own theoretical commitments, setting the scene for the presentation of our own model.

90 2.1 Metaphor processing and comprehension in human participants

91 Behavioral and electrophysiological research with human participants has gone a long way in clarifying
92 the cognitive mechanisms involved in metaphoric language processing and comprehension. In most
93 behavioral studies, participants decide whether literal and metaphoric sentences make sense (a semantic
94 judgement task), while the reaction times and accuracy are measured and compared across the different
95 sentence types. In electrophysiological studies, in addition to the behavioral data, Event-Related Potentials
96 (ERP) are analysed. ERPs are brain responses to specific cognitive events, in this case to literal and
97 metaphoric sentences presented to the participants. Both behavioral and ERP studies on metaphor

98 processing have shown that metaphor processing and comprehension are modulated by the conventionality
99 level of metaphoric utterances.

100 Analyses of behavioral data obtained from participants in response to literal and metaphoric utterances
101 have revealed longer reaction times and lower accuracy rates when participants judge novel metaphors
102 than literal sentences. Conventional metaphoric sentences evoke either shorter reaction times than novel
103 metaphoric, but longer than literal sentences (Lai and Curran, 2013), or comparable reaction times to
104 literal items (Arzouan et al., 2007). In electrophysiological research, two ERP components have garnered
105 particular interest in this line of work. The N400, a negative-going wave elicited between 300-500ms post-
106 stimulus, was first reported in response to semantic anomaly (Kutas and Hillyard, 1984), with meaningless
107 sentences evoking larger N400 amplitudes than meaningful sentences. In line with previous suggestions
108 and a recently proposed single-stream Retrieval-Integration account of language processing, the N400
109 can be interpreted as reflecting retrieval of information from semantic memory (Brouwer and Hoeks,
110 2013; Brouwer et al., 2017; Kutas and Federmeier, 2000). Other accounts propose that the N400 can
111 be seen as reflecting both information retrieval and integration (Coulson and Van Petten, 2002; Lai and
112 Curran, 2013). In electrophysiological research on metaphor, novel metaphors evoke larger N400 amplitudes
113 than conventional metaphors, followed by literal utterances, which evoke the smallest N400 amplitudes
114 (Arzouan et al., 2007). This graded effect might reflect an increase in retrieval of semantic information
115 required for complex mappings in the case of metaphoric utterances, which is additionally modulated by
116 the conventionality of the metaphor.

117 Another ERP component that has recently received attention in the context of metaphor comprehension
118 is the late positive complex (LPC). LPC is a positive-going wave observed between 500 and 800ms
119 post-stimulus. While LPC amplitudes observed in response to conventional metaphors converge with those
120 for literal utterances, novel metaphors evoke reduced LPC amplitudes (Arzouan et al., 2007; Bambini
121 et al., 2019; Goldstein et al., 2012; Rataj et al., 2018). This reduction is difficult to interpret within the
122 current theories of the LPC, which see this component as reflecting integration of the retrieved semantic
123 information in a given context. Because semantic integration demands are larger for novel metaphoric than
124 literal sentences, as evident in behavioral data, larger LPC amplitudes for novel metaphoric than literal
125 sentences would be expected. Such increases in LPC amplitudes have been reported in studies that used
126 conventional metaphors, or metaphors that were evaluated as neither familiar nor unfamiliar (De Grauwe
127 et al., 2010; Weiland et al., 2014), but not when the tested metaphoric utterances were novel. One possible
128 interpretation of this novel metaphor effect is that because of the difficulty related to establishing novel
129 mappings in the course of novel metaphor processing, access to semantic information that begins in the
130 N400 time window is prolonged and reflected in sustained negativity that overlaps with the LPC, thus
131 reducing its amplitude. Taken together, ERP findings reveal crucial information about the time-course of
132 metaphor processing and comprehension, and point to two cognitive mechanisms, i.e., semantic information
133 retrieval and integration, as the core operations required in understanding metaphoric language.

134 Several theoretical accounts of metaphor processing and comprehension have been formulated.
135 The *structure mapping model* (Bowdle and Gentner, 2005; Wolff and Gentner, 2011) proposes that
136 understanding metaphoric utterances such as *this classroom is a zoo* require a symmetrical mapping
137 mechanism to align relational commonalities between the source (*zoo*) and target (*classroom*), as well
138 as an asymmetrical mechanism projecting an inference about the source to the target. The *career of*
139 *metaphor model* (Bowdle and Gentner, 2005) further posits that conventional metaphor comprehension
140 requires a process of categorization, while novel metaphors are understood by means of comparison. Within
141 the conceptual expansion account, the existing concepts are broadened as a results of novel meaning

142 construction (Rutter et al., 2012; Ward, 1994). Conceptual expansion could be seen as creating a re-
143 representation of an existing concept in the process of novel meaning construction. The important questions
144 thus concern the ways the semantic knowledge is retrieved and integrated in the process of metaphoric
145 meaning construction.

146 2.2 Computational studies

147 From the perspective of semantic representation, computational approaches to modelling metaphor have
148 typically sought some mechanism for identifying the transference of salient properties from one conceptual
149 domain to another (Shutova, 2015). Some approaches have used structured, logical representations: one
150 early exemplar is the MIDAS system of Martin (1990), which maps metaphors as connections between
151 different conceptual representations, interpreting the semantic import of a metaphor in terms of plausible
152 projections of properties from one concept to another. The system described by Narayanan (1999)
153 likewise builds up conceptual representations as composites of properties, introducing a concept of broader
154 conceptual domains grounded in knowledge about action in the world which can be mapped to one another
155 by identifying isomorphisms in patterns of relationships within each domain. This move opens up a
156 correspondence between computational methodologies and the theory of *conceptual metaphor* outlined by
157 Lakoff and Johnson (1980). Barnden (2008) offers an overview of these and a few other early approaches,
158 tying them in to the rich history of theoretical and philosophical work on metaphor.

159 Data-driven approaches have often adopted a similar theoretical premise to metaphor (seeking to model
160 cross-domain mappings), but build representations based on observations across large-scale datasets
161 rather than rules or logical structures. So, for instance, the model developed by Kintsch (2000) extracts
162 statistics about dependency relationships between predicates and subjects from a large-scale corpus and
163 then iteratively moves from a metaphoric phrase to a propositional interpretation of this phrase by traversing
164 the relationships implied by these statistics. Similarly, Utsumi (2011) uses co-occurrence statistics to build
165 up representations, pushing labelled word-vectors into a *semantic space* in which geometric relationships
166 can be mapped to predictions about word meaning: proximity between word-vectors in such a space are
167 used to generate plausible interpretations of metaphors. Shutova et al. (2012a) present a comprehensive
168 review of statistical approaches to the computational modelling of metaphor.

169 A recent development in these approaches (and in natural language processing in general) has been
170 the application *distributional semantic* techniques to capture phrase and sentence level semantics via the
171 geometry of vector spaces. The distributional semantic paradigm has its roots in the theoretical work of
172 Harris (1957), and particularly the premise that words that tend to be observed with similar co-occurrence
173 profiles across large scale corpora are likely to be related in meaning; modern computational approaches
174 capture this by modelling words as vectors in high-dimensional spaces which capture the details of those
175 co-occurrence profiles. Features of these vectors and spaces have been shown to improve performance in
176 natural language processing tasks ranging from word sense disambiguation (Schütze, 1998; Kartsaklis
177 and Sadrzadeh, 2013) and semantic similarity ratings (Hill et al., 2015) to more conceptually structured
178 problems such as analogy completion (Mikolov et al., 2013; Pennington et al., 2014).

179 A preponderance of computational schemes for traversing corpora and generating mathematically
180 tractable vector-space representations have been developed (see Clark, 2015, for a fairly recent and
181 inclusive survey). However, the basic insight can be captured by imagining a large matrix in which each
182 row is a vector corresponding to a word in our vocabulary. The columns of this matrix — the *co-occurrence*
183 *dimensions* — correspond to words which have been observed co-occurring with a vocabulary word. The
184 value of the entry at row w and column c represents the probability of observing vocabulary word w in

185 the context of *c*. Words with similar meanings have similar co-occurrence profiles, and thus similar row
186 vectors, and this similarity can now be measured in mathematical terms. Many variants exist: matrix values
187 are often chosen not as raw probabilities but *pointwise mutual information* values (normalising the raw
188 probabilities for those expected due to the words' overall frequency); matrices are often factorised to reduce
189 dimensionality and smooth the estimates, or learned using neural networks rather than direct statistics
190 (Mikolov et al., 2013). Co-occurrence can be defined at the level of sentence or whole documents, of words
191 or characters, or in terms of syntactic dependency or other semantic relations (Schütze, 1992; Padó and
192 Lapata, 2007; Kiela and Clark, 2014; Levy and Goldberg, 2014a); although it is usually taken as simple
193 lexical co-occurrence within a fixed-width window of words within sentences. Even this simple version can
194 vary in terms of the co-occurrence window width, with some evidence that the slide from small to large
195 co-occurrence windows might correspond to shifts along semantic spectra such as that of concreteness to
196 abstractness (Hill et al., 2013).

197 In terms of modelling metaphor, distributional semantic models have been used to generate contextually
198 informed paraphrases of metaphors (Shutova et al., 2012b), have played a role as components in more
199 complex classifiers (Tsvetkov et al., 2014), and have even been used to interface between linguistic and
200 visual data (Shutova et al., 2016). The linear algebraic structure of distributional semantic representations
201 lends itself to composition, in that mathematical operations between word-vectors can be mapped to
202 sequences of words, and interpretations of larger linguistic compositions can therefore potentially be
203 pushed into a computational model (Coecke et al., 2011). Gutiérrez et al. (2016) have exploited this aspect
204 of high-dimensional semantic representations to model metaphoric adjective-noun phrases as operations
205 between a vector (representing a noun) and a second-order tensor (representing an adjective), by which
206 the adjective-tensor projects the noun-vector into a new region of a semantic space. So, for instance,
207 *brilliant child* is represented by a composed vector that we might expect to find in the vicinity of words
208 like *intelligent* rather than words like *glowing*.

209 2.3 The Role of Context

210 These approaches, however, give little attention to the role of *gradedness* and *context* in the processing of
211 metaphor; but many theoretical approaches point out that these play a vital role. The relevance-theoretic
212 *deflationary account* of Sperber and Wilson (2008), for example, proposes that metaphor can be understood
213 as occupying a region within a spectrum (or perhaps more properly, a region in a multi-dimensional
214 landscape) of various linguistic phenomena that come about in the course of communication. Metaphoricity
215 thus exists not as a binary distinction but on a scale, and as part of a larger scale (and we will see this
216 reflected the data described in Section 4 below).

217 Carston (2010b) emphasises context-specificity: she argues that there are two different modes of metaphor
218 processing, and that what might be thought of as the more basic and on-line mode involves the construction
219 of *ad hoc* concepts. So, to process a metaphoric verb-object phrases such as *murder wonder*, an ephemeral
220 concept of an activity MURDER* has to be formulated on the spot, and in the context of the application
221 of the phrase. Furthermore, the propositional content of the phrase, to the extent we embrace the idea that
222 language is propositional, begins to become blurred as components of imagery and phenomenology begin
223 to infiltrate language. The idea that metaphoric language involves an extemporaneous projection of a new
224 conceptual framework presents a challenge to cognitivist approaches to metaphor, typified by the theory of
225 conceptual metaphors (Lakoff and Johnson, 1980; Gibbs and Tendahl, 2006), in that it requires a capacity
226 for the construction of *ad hoc* spaces of lexical semantic representations susceptible to the influences of a
227 complex and unfolding situation in which communication between cognitive agents is happening.

228 This approach therefore questions the idea that metaphor involves mappings between established concepts.
229 To take an example from the data we will model below, the conventional metaphor *cut pollution* arguably
230 involves the construction of an *ad hoc* concept CUT*, which extends the action denoted by the verb to
231 something that can be done to *pollution*, in line with Carston (2010a). This is in contrast to a cognitive
232 linguistic perspective on metaphor, which would seek to find a sense in which a fixed property of CUTTING
233 is transferred to the object *pollution*. In the next sections, we show how a computational method can be
234 developed which follows the *ad hoc* concept view, and test its ability to model human judgements.

3 COMPUTATIONAL METHODOLOGY

235 With a sense of the way that metaphor fits into a broader range of human semantic representations, we
236 now turn to the task of modelling metaphor computationally. Our objective here is to explore whether and
237 how we can apply statistical analysis of large-scale language corpus data to the problem of re-representing
238 metaphor. Working from the theoretical premise that metaphor emerges in a particular semantic context,
239 we use a methodology for systematically generating on-line lexical semantic relationships on the basis of
240 contextualising information.

241 3.1 Approach

242 Our approach is based in the standard distributional semantic view of geometric semantic representation:
243 construction of word meanings as vectors or points that are meaningful in terms of their relationship to one
244 another in some appropriate space, defined in terms of word co-occurrence statistics across a large scale
245 corpus. The distinctive feature of our approach, though, is that the semantic re-representation associated
246 with metaphor interpretation will be expressed as projection into a series of geometric subspaces, each
247 determined in an on-line way on the basis of context. Our model, then, like that of Gutiérrez et al. (2016),
248 seeks to represent metaphor in terms of projections in geometric spaces; however, rather than simply
249 use linear algebraic operations to move or compare word representations within a single static space, we
250 propose to model every instance of a metaphoric composition in terms of a newly generated subspace,
251 specific to the conceptual context in which the metaphor occurs.

252 This subspace is based on a particular composition (in the experiments below, a two-word verb-noun
253 phrase, but the method is general): its dimensions are chosen as the most salient features — the strongest
254 statistical co-occurrence associations — which the words in the phrase have in common. It is thus distinct
255 in its geometry from the space which would be defined for other compositions using one or the other but
256 not both words. We hypothesize that these dimensions will provide us both an appropriate mechanism for
257 specifying *ad hoc* contextualised projections, and adequate measures for modelling the dynamic production
258 of semantic representations; we test this by learning statistical models based on the geometric properties
259 of the subspaces and the relative positioning of the words within them, and evaluating their ability to
260 predict the metaphoricity of the compositional phrases. To be clear, our objective is not to refute the
261 cognitive stance on metaphor; rather, we seek to provide a methodology that accommodates a pragmatic
262 interpretation of metaphor as a means for communication about extemporaneously constructed concepts,
263 an objective that has proved elusive for computational models.

264 This context-dependent modelling approach was originally developed by Agres et al. (2015), and further
265 developed by McGregor et al. (2015), for the purposes of context-dependent concept discovery. McGregor
266 et al. (2017) showed that a variant could provide a model of the phenomenon of semantic type coercion of
267 the arguments of verbs in sentential context; and Agres et al. (2016) showed that distances in the contextual
268 subspaces were more closely associated with human judgements of metaphoricity than distances in standard

269 static distributional semantic models. Here, our hypothesis is that this can be used to provide a model of
 270 metaphor more generally: that the on-line projection of context specific conceptual subspaces can capture
 271 the process of re-representation inherent in the construction of the *ad hoc* concepts necessary to resolve the
 272 semantics of a non-literal phrase.

273 3.2 Data Cleaning and Matrix Building

274 In order to select subspaces suitable for the geometric analysis of word-pairs in the context of a set of
 275 co-occurrence dimensions, we begin by building a *base space* from co-occurrence statics over a large
 276 textual corpus, using standard distributional semantic techniques. We use the English language component
 277 of Wikipedia, and begin by applying a data cleaning process which removes punctuation (aside from
 278 apostrophes and hyphens), converts all text into lower case, and detects sentence boundaries. The resulting
 279 corpus consists of almost 1.9 billion word tokens representing about 9 million word types, spread across
 280 just over 87 million sentences.

281 We consider the 200,000 most frequent word types in the corpus to be our vocabulary, and our base
 282 space will accordingly be a matrix consisting of 200,000 rows (vocabulary word types) and some 9 million
 283 columns (co-occurrence word types). We use the standard approach of defining co-occurrence simply as
 284 observation within a fixed window within a sentence; here we use a symmetric window of 2x2 words.
 285 While broader windows have been reported as being suited for capturing specific semantic properties,
 286 small windows have proved particularly good for modelling general semantic relatedness; as we are
 287 seeking to analyse the paradigmatic relationships inherent in distributional semantics, rather than the type
 288 of syntagmatic relationships that emerge over a larger number of words, we choose to focus on smaller
 289 co-occurrence windows here (Sahlgren, 2008).

290 For the matrix values we use a variant of pointwise mutual information (PMI): given a vocabulary word w
 291 and a word c observed co-occurring with w , a frequency of observed co-occurrences $f(w, c)$, independent
 292 frequencies of $f(w)$ and $f(c)$ respectively, and a total count of vocabulary word occurrences W , we define
 293 the mutual information between w and c as follows:

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

294 Here a is a smoothing constant applied to weight against the selection of very infrequent dimensions in the
 295 contextual projection procedure that will be described below. This value is set to 10,000, based on trial and
 296 error, but this value also turns out to be roughly equal to the mean frequency of all co-occurrence words,
 297 meaning that the average ratio of frequencies will be approximately halved; PMI values associated with
 298 very rare co-occurrence terms will be severely punished, while values for very common co-occurrence
 299 terms will be relatively unaffected. The addition of 1 to the ratio of frequencies guarantees that all PMI
 300 values will be non-negative, with a value of 0 indicating that the words w and c never co-occur with one
 301 another. It should be noted that this expression is approximately equivalent to the logarithm of the ratio of
 302 the joint probability of w and c co-occurring, skewed by the smoothing constant and the incrementation of
 303 the ratio.

304 This PMI equation is similar to established methods for weighting co-occurrence statistics, but differs in
 305 some important ways that are designed to accommodate the contextual and geometric objectives of our
 306 own methodology. In a standard statistical approach to distributional semantics, the information theoretical
 307 insight of a PMI type measure is that frequent observations of co-occurrences with infrequent words should

308 be given heavily positive weightings. That idea holds for our own approach up to a point, but, as we would
309 like a mechanism for selecting co-occurrence features that are conceptually salient to multiple words,
310 we would like to avoid giving preference to co-occurrence terms that are so infrequent as to be virtually
311 exclusive to a single word or phrase. Adding a balances the propensity for distributional semantic models
312 to emphasise extremely unlikely observations, as this factor will have less of an impact on terms that
313 already have a relatively high overall frequency $f(c)$. By guaranteeing that all our features are non-negative,
314 we can reliably project our word-vectors into contextualised subspaces characterised by not only angular
315 relationships between the word-vectors themselves, but also with a more informative geometry including a
316 sense of extent, centre, and periphery. The merits of this approach will be discussed further in Section 3.4.

317 3.3 Projecting Contextualised Subspaces

318 The procedure described in Section 3.2 results in a large and highly informative but also sparse matrix
319 of co-occurrence information, where every observed co-occurrence tendency for all the words in our
320 vocabulary is systematically tabulated. To give a sense of the scope of this representational scheme, every
321 one of the 9 million word types that come up in our corpus becomes the label of a co-occurrence dimensions,
322 but the distribution of word frequencies is characterised by the long tail familiar to corpus linguists, with
323 5.4 million of the 9 million word types in the corpus co-occurring with one of the 200,000 vocabulary
324 words 10 times or less.

325 Our next task is to establish a set of techniques for extrapolating *ad hoc* representations capturing the
326 contextualisation of the semantics associated with a particular denotation, something that is crucial to
327 metaphoric re-representation. The premise we will work off of is the distributional hypothesis, namely,
328 that consistencies in co-occurrence between two lexical semantic representations correspond to semantic
329 relatedness between the words being represented. Building off of this idea, we propose that there should
330 be subsets of co-occurrence dimensions which are salient to particular conceptual contexts. Given the
331 looseness and ambiguity inherent in word use, and the relationship between this and the drift from literal to
332 figurative language, we suggest that there are groups of co-occurrence dimensions that can collectively
333 represent either observed or potential contexts in which a word can take on particular semantic aspects.

334 Consider the sets of co-occurrence terms with the highest average PMI values for the words *brilliant*
335 *diamond* and *brilliant child*, the first of which is likely to be interpreted as a literal phrase, the second of
336 which is a metaphor, albeit a conventionalised one:

- 337 1. **brilliant diamond** *carat, koh-i-noor, carats, diamonds, diamond, emerald, barbra, necklace, earrings,*
338 *rose-cut*
- 339 2. **brilliant child** *prodigy, precocious, prodigies, molestation, sickly, couple's, destiny's, intellectually,*
340 *unborn, imaginative*

341 Here we can see how the alteration in the noun modified by *brilliant* skews the profile of co-occurrence
342 terms with the highest joint mean into two different conceptual spaces. For the literal phrase *brilliant*
343 *diamond*, we see co-occurrence terms which seem logically associated with denotations and descriptions of
344 gems, such as *emerald* and *carat*, as well as applications such as *earrings* and specifications such as *rose-cut*.
345 In the case of *brilliant child*, on the other hand, we see words which could stand in as interpretations of the
346 metaphor *brilliant*, such as *prodigy*, or, perhaps with some licence, *precocious*, as well as terms related
347 generally to children.

348 In both cases we also note some unexpected terms creeping in. In the case of *brilliant child*, an analysis of
349 the corpus suggests that the inclusion of *destiny's* is a reference to the music group *Destiny's Child*, who are

350 sometimes described by critics cited in our corpus as “brilliant”. A similar analysis of co-occurrences of the
 351 name *Barbra* with *brilliant* and *diamond* across Wikipedia reveals that Barbra Streisand has periodically
 352 performed with Neil Diamond, and that she is another artist who has often been acclaimed as “brilliant”.
 353 These co-occurrences offer up instances of how elements of ambiguity can enter into relationships between
 354 distributional semantic representations: while there is always an explanation for the presence of such
 355 dimensions in this type of analysis, there is not an interpretation that is particularly coherent conceptually.

356 One of the strengths of distributional semantic models, though, is that the high-dimensional spaces
 357 they inhabit tend to be fairly resilient against noise. This propensity for using dimensionality to support
 358 representations that are, overall, semantically apt aligns with our hypothesis that there should be subsets of
 359 dimensions which, taken collectively, represent conceptual contexts. We would like to develop a model
 360 which allows for the systematic selection of subspaces of co-occurrence dimensions, based on input
 361 consisting of individual words, which on the whole capture something of the conceptual context in which
 362 these terms might be composed into a phrase. These techniques, we propose, will allow us to project
 363 re-representations of the lexical items involved in the phrase that will facilitate the analysis of how their
 364 semantics could metaphorically interact.

365 With this in mind, we propose to explore three different techniques for selecting subspaces based on an
 366 analysis of the co-occurrence profiles of two different input words:

- 367 1. MEAN: We take the co-occurrence terms with the highest arithmetic mean PMI value across input
 368 words;
- 369 2. GEOM: We take the co-occurrence terms with the highest geometric mean PMI value across input
 370 words;
- 371 3. INDY: We take a concatenation of the co-occurrence terms with the highest PMI values for each word
 372 independently.

373 For the MEAN technique, given two input words w_1 and w_2 , the value for any candidate co-occurrence term
 374 c_j is simply:

$$M(c) = (PMI(w_1, c_j) + PMI(w_2, c_j))/2$$

375 We can take the value for every co-occurrence term and then select the top k such terms and project our
 376 input words into the corresponding space. For the GEOM technique, we similarly apply the equation for the
 377 geometric mean of PMI values:

$$G(c_j) = \sqrt{PMI(w_1, c_j) \times PMI(w_2, c_j)}$$

378 Here it should be noted that, while this equation is strictly defined to include PMI values of 0, the outputs
 379 for any such terms would be 0, and so we are in practice only interested in co-occurrence terms with
 380 non-zero PMI values for both input words. There is not a rational definition for the geometric mean of a
 381 set of inputs containing negative numbers, but, returning to Equation 1 above, we recall that our matrix
 382 contains only non-negative elements, anyway.

383 For the INDY technique, we apply an additional constraint to avoid selecting a co-occurrence term that
 384 has a high PMI value for both input terms twice. We iteratively select the co-occurrence term with the top
 385 PMI value for each input, and, if we encounter a term for one input that was already selected for the other
 386 input, we move to the next highest scoring term that hasn't already been selected. We carry this process on
 387 until we have established a subspace with k dimensions.

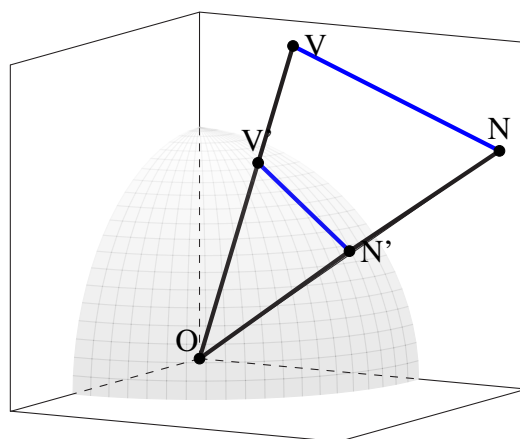


Figure 1 Two word-vectors projected into a contextualised subspace, and the unit sphere intersecting the normalised version of each vector.

388 The final parameter of this component of our model is k itself, the dimensionality of the subspaces
 389 selected using any of the techniques now defined. For the purpose of experiments reported here, we will
 390 use a value of 200. This value is low enough to guarantee that we can define spaces for the GEOM technique
 391 that involve dimensions with non-zero values for both input words, but on the other hand large enough
 392 to hopefully build subspaces that are robust against noise and capture some of the conceptual nuance
 393 inherent in the interaction between the input terms as a composed phrase. Other values for k have been
 394 explored elsewhere (McGregor et al., 2015, 2017), and 200 has generally returned good results. In the
 395 present work, our objective is to focus on the alignment of our methodology with theoretical stances on
 396 semantic re-representation; there is clearly room for further exploration of the model's parameter space in
 397 future work.

398 An example of a subspace with two word-vectors projected into it is illustrated in Figure 1. Some of the
 399 primary element of such a space are also indicated here: in addition to the distance from the origin of each
 400 of the word-vectors (represented by the points V and N), the distance between the vectors \overline{VN} is also an
 401 essential measure of the semantic relationship between the two words labelling these vectors, indicating
 402 the degree of overlap between these words in the context of the projection they jointly select. Furthermore,
 403 a standard technique in distributional semantics is to consider the normalised vectors. To this end, a unit
 404 sphere intersecting the vectors is illustrated, and we note that the distance between the normalised vectors
 405 V' and N' correlates monotonically with the angle $\angle VON$. These will now serve as a basis for a much
 406 more involved analysis of the statistical geometry of a contextualised subspace.

407 3.4 Geometric Analysis of Contextualised Projections

408 The techniques for analysing co-occurrence terms associated with potentially metaphoric phrases
 409 described in the previous section result in the projection of subspaces in which the word-vectors
 410 corresponding to the input words, and for that matter any other word-vector in our base space, maintain a
 411 fully geometric aspect. The dimensions of the subspace are labelled by the co-occurrence terms selected,
 412 and the values for a word-vector along these dimensions are simply specified by the corresponding value in
 413 the full base space.

414 Because our base space is not normalised, there is, for any word-vector, a notion of distance from the
 415 origin of a subspace: the value for any given coordinate of word-vector w_i for co-occurrence dimension d_j
 416 will be $PMI(w_i, d_j)$, which could range from 0 if the word never co-occurs with that term to something

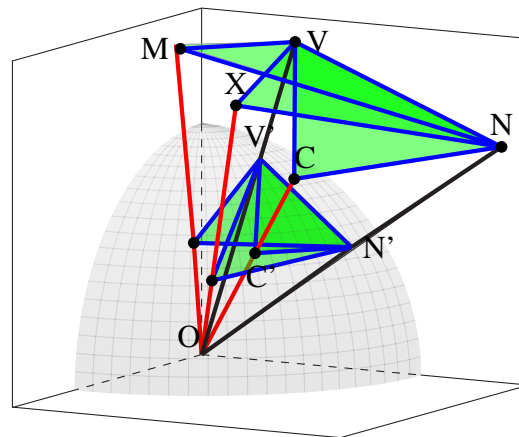


Figure 2 The geometry of a contextually projected subspace. V and N are verb and noun vectors, while M , X , and C are the mean, maximum, and central vectors. V' , N' , M' , X' , and C' are their norms, where they intersect the unit sphere.

417 quite large if the word is on the one hand frequent and on the other hand often co-occurs with a term that is
 418 similarly frequent. So, in a given subspace, if a particular word has high PMI values across a number of the
 419 co-occurrence dimensions, we would expect it to be far from the origin. Conversely, a word with mainly
 420 low and zero PMI values would be close to the origin.

421 Furthermore, because our subspaces consist only of elements with non-negative values, there is a sense of
 422 centre and periphery to them. So, for instance, a word-vector with high PMI values for a few co-occurrence
 423 dimensions in a given space but low values for most of the dimensions would be skewed away from the
 424 centre. On the other hand, a word-vector with consistent values across dimensions would be relatively close
 425 to the centre of the space (though not far from the origin if these values were consistently low).

426 Word-vectors will naturally have relationships to one another, as well. There is a Euclidean distance
 427 between them, an angle between them, and relative distances from the origin. There will also be a number
 428 of what we will term *generic vectors* in the space, meaning points corresponding to values characteristic of
 429 the space overall rather than any particular word-vector projected into that space. In particular, we define a
 430 *mean-vector*, where each element of the vector is the mean value of all word-vectors with non-zero values
 431 for each corresponding co-occurrence dimension, a *maximum-vector*, where each element is the highest
 432 value for any word-vector along each corresponding dimension, and a *central-vector*, which is simply a
 433 uniform vector in which each element is the mean of the mean-vector.

434 We suggest that these geometric features provide a basis for an analysis of the way in which co-occurrence
 435 observations across a large-scale corpus can map to information about metaphoricality and attendant re-
 436 representation. In addition to properties such as centrality within the space and distance from the origin
 437 discussed above, the relationship between two word-vectors relative to a central or maximal point in a
 438 subspace should tell us something about the way that they interact with one another semantically: words
 439 with similarly lopsided co-occurrence profiles within a subspace will be skewed in the same direction,
 440 for instance, and so may be expected to share an affinity within the conceptual context being modelled.
 441 Relative distances from generic vectors and also from the origin might also be expected to predict semantic
 442 relationships between words. And finally, the characteristics of the space itself, potentially inherent in
 443 the generic vectors and their interrelationships outside any analysis of actual word-vectors, might tell us
 444 something about the underlying context of the generation of the space in the first place.

Table 1 List of measures for geometric analysis of subspaces, with reference to Figure 2 .

	FULL VECTORS	NORMALISED VECTORS
<i>distances</i>	$\overline{V}, \overline{N}, \overline{VN}, \overline{M}, \overline{X}, \overline{C}$	$\overline{V'N'}$
<i>means</i>	$\mu(\overline{VM}, \overline{NM}), \mu(\overline{VX}, \overline{NX}), \mu(\overline{VC}, \overline{NC})$	$\mu(\overline{V'M'}, \overline{N'M'}), \mu(\overline{V'X'}, \overline{N'X'}),$ $\mu(\overline{V'C'}, \overline{N'C'})$
<i>ratios</i>	$(\overline{VM} : \overline{NM}), (\overline{VX} : \overline{NX}), (\overline{VC} : \overline{NC})$	$(\overline{V'M'} : \overline{N'M'}), (\overline{V'X'} : \overline{N'X'}),$ $(\overline{V'C'} : \overline{N'C'})$
<i>fractions</i>	$\overline{V}/\overline{N}, \overline{VM}/\overline{NM}, \overline{VX}/\overline{NX}, \overline{VC}/\overline{NC},$ $\mu(\overline{V}, \overline{N})/\overline{M}, \mu(\overline{V}, \overline{N})/\overline{X}, \mu(\overline{V}, \overline{N})/\overline{C},$ $\overline{C}/\overline{M}, \overline{C}/\overline{X}, \overline{M}/\overline{X}$	$\overline{V'M'}/\overline{N'M'}, \overline{V'X'}/\overline{N'X'}, \overline{V'C'}/\overline{N'C'}$
<i>angles</i>	$\angle VON, \angle VMN, \angle VXN, \angle VCN,$ $\angle MOC, \angle MOX, \angle COX$	$\angle V'M'N', \angle V'X'N', \angle V'C'N'$
<i>areas</i>	$\triangle VMN, \triangle VXM, \triangle VCM$	$\triangle V'M'N', \triangle V'X'M', \triangle V'C'M'$

445 Figure 2 illustrates a subspace with all its characteristic features: the word vectors *V* and *N* which
 446 generate and then are subsequently projected into the subspace along with the mean, maximum, and central
 447 vectors, and then the various relationships which we propose to analyse in the context of metaphoricity. (*V*
 448 and *N* stand for *verb* and *noun*; as will be seen in Section 4, the input to our space will be the components
 449 of potentially metaphoric verb-object phrases.) In addition to the aforementioned vectors, we also consider
 450 the normalised versions of each these vectors, which should provide us with a basis for considering the
 451 centrality of word-vectors. For instance, a verb-vector and noun-vector might have quite different lengths,
 452 and so could potentially form an obtuse angle with the mean-vector as a vertex ($\angle VMN$), but they might
 453 both be to the same side of *M* in the space and so form an acute angle on a unit sphere ($\angle V'M'N'$).

454 We define a total of 48 geometric features in any given subspace. These encompass distances, means of
 455 distances, ratios of distances, angles, areas of triangles defined by distances, and a number of these features
 456 taken at the surface of the hypersphere representing normalisation of vectors. They are itemised in Table 1.
 457 Distances comprise the norms of vectors and the Euclidean distances between vectors, while means are the
 458 averages of some pairs of these distances. Ratios involve the fraction of the lower of a pair of distances
 459 over the higher, and are intended to provide a comparative measure of the relationship between vectors
 460 without presuming one as the numerator and the other as the denominator of a fraction. Fractions do take
 461 one vector norm or one mean of vector norms as an absolute denominator. Angles are taken both at the
 462 origin and at the vertices of generic vectors, and areas measure the triangles indicated by a subset of these
 463 angles.

464 Collectively, these measures describe all the components of the geometry of a contextualised distributional
 465 semantic subspace which we will explore for indications of metaphoric re-representation. In the experiments
 466 described in Section 5, they will become the independent variables defining a set of models that will
 467 seek to learn to predict metaphoricity, meaningfulness, and familiarity in verb-object phrases. They will
 468 likewise serve as tools for interpreting the behaviour of these models: the ability to trace these features back
 469 to co-occurrence phenomena will prove to be a useful mechanism for understanding the ways in which
 470 statistics derived from a large collection of text can be mapped to semantic phenomena associated with the
 471 contextualisation inherent in conceptualisation.

472 3.5 Establishing a Baseline

473 In order to compare our dynamically contextual distributional semantic methodology, which has been
474 specifically designed to capture the way that re-representation occurs in a cognitive and environmental
475 context, with more standard distributional semantic techniques, we model our data using the word-vectors
476 output by the widely reported `word2vec` methodology (Mikolov et al., 2013). This approach involves
477 building a neural network which learns word-vectors by iteratively observing the ways that words co-occur
478 in a corpus. The algorithm begins by randomly assigning each word in its vocabulary a word-vector in
479 a normalised vector space, and then, each time a word is observed in a particular context, it adjusts the
480 values of the corresponding word-vector slightly to pull it towards vectors corresponding to words observed
481 in similar contexts.

482 The `word2vec` technique is different from our dynamically contextual approach in two important ways.
483 First of all, it projects word-vectors into a normalised hypersphere of arbitrary dimensionality, meaning
484 that the only measure for comparing two lexical semantic representations to one another is cosine similarity
485 (which will correlate monotonically with Euclidean distance in a normalised space). This means that there
486 is no mechanism for extracting the wider range of geometric features we use to examine the nuances of
487 semantic phenomena, such as distance from origin, centrality, or relation to generic vectors.

488 Second, and perhaps even more importantly, because the word-vectors learned by a neural network are
489 *abstract* in the sense that their dimensions are just arbitrary handles for making slight adjustments to
490 relationships between vectors, there is no way to meaningfully select dimensions for the projections of
491 lower dimensional subspaces corresponding to particular conceptual contexts. In fact, Levy and Goldberg
492 (2014b) make a compelling case for considering this approach as being commensurate with the matrix
493 factorisation techniques for building semantic representations described by Deerwester et al. (1990),
494 enhanced with a large number of modelling parameters.

495 We build a `word2vec` model based on the same corpus described in Section 3.2, applying the *contextual*
496 *bag-of-words* procedure outlined by Mikolov et al. (2013) to generate a 200 dimensional vector space based
497 on observations within a 2x2 word co-occurrence window.¹ This model will serve as a point of comparison
498 with our own dynamically contextual distributional semantic methodology, offering up a singular space in
499 which lexical semantic representations are simply compared in terms of their universal relationship to one
500 another, without any mechanism for generating *ad hoc* relationships in a contextually informed way.

501 4 HUMAN METAPHOR JUDGEMENTS

502 In this study, we seek to develop a computational model of the way that metaphor emerges in a particular
503 conceptual context, as a linguistic artefact situationally endowed with an unfamiliar meaning. Our empirical
504 objective will be to predict the extent to which multi-word phrases would be perceived as metaphoric.
505 In order to generate data for this modelling objective, and also to understand the relationship between
506 metaphor and other semantic categories, we introduce a dataset of verb-object compositions evaluated by
507 human judges, and perform some preliminary analyses on correlations between the human judgements.

507 4.1 Materials

508 The materials are verb-noun word dyads, which were originally selected for an ERP study on metaphor
509 comprehension in bilinguals (Jankowiak et al., 2017). Five normative studies were performed prior to the

¹ This is implemented using the Gensim module for Python.

Normative study type	Number of participants(female)	Mean age
Cloze probability	140 (65)	23
Meaningfulness ratings	133 (61)	22
Familiarity ratings	101 (55)	23
Metaphoricity ratings	102 (59)	22

Table 2 Demographic characteristics of participants of the four normative studies, including the number of participants (number of female participants) and mean age.

510 ERP experiment to confirm that the word pairs fell within the following three categories: novel metaphors
 511 (e.g., *to harvest courage*), conventional metaphors (e.g., *to gather courage*), and literal expressions (e.g., *to*
 512 *experience courage*). Based on the results of the normative studies, the final set of 228 English verb-noun
 513 word dyads (76 in each category) was selected for the purpose of the current study. The main results of
 514 the four normative studies performed prior to the EEG study will be reported here; for a more detailed
 515 discussion of the materials see Jankowiak et al. (2017). Mixed-design analyses of variance (ANOVAs) with
 516 utterance type as a within-subject factor and survey block as a between-subject factor were conducted.
 517 There was no significant main effect of block. Significance values for the pairwise comparisons were
 518 corrected for multiple comparisons using the Bonferroni correction. The Greenhouse-Geisser correction
 519 was applied whenever Mauchly's test revealed the violation of the assumption of sphericity, and in these
 520 cases, the original degrees of freedom are reported with the corrected p value.

521 4.1.1 Cloze probability

522 To ensure that expectancy effects caused by participants anticipating the second word in a given word
 523 dyad would not impact the results of the EEG study, a cloze probability test was performed. Participants
 524 received the first word of a given word pair, and provided the second word, so that the two words would
 525 make a meaningful expression. If a given word pair was observed more than 3 times in the cloze probability
 526 test, the word dyad was excluded from the final set and replaced with a new one. This procedure was
 527 repeated until the mean cloze probability for word pairs in all four conditions did not exceed 8% (novel
 528 metaphoric, conventional metaphoric, and meaningless word pairs ($M = 0, SD = 0$); literal word pairs
 529 ($M = .64, SD = 2.97$)).

530 4.1.2 Meaningfulness

531 Participants of this normative test rated how meaningful a given word pair was on a scale from 1
 532 (totally meaningless) to 7 (totally meaningful). A main effect of utterance type was found, [$F(3, 387) =$
 533 $1611.54, p < .001, \epsilon = .799, \eta_p^2 = .93$]. Pairwise comparisons showed that literal word pairs were
 534 evaluated as more meaningful ($M = 5.99, SE = .05$) than conventional metaphors ($M = 5.17, SE = .06$)
 535 ($p < .001$), and conventional metaphors as more meaningful than novel metaphors ($M = 4.09, SE =$
 536 $.08$)($p < .001$).

537 4.1.3 Familiarity

538 Familiarity of each word pair was assessed in another normative study, in which participants decided how
 539 often they had encountered the presented word pairs on a scale from 1 (very rarely) to 7 (very frequently).
 540 A main effect of utterance type was found, [$F(2, 296) = 470.97, p < .001, \epsilon = .801, \eta_p^2 = .83$]. Pairwise
 541 comparisons showed that novel metaphors ($M = 2.15, SE = .07$) were rated as less familiar than
 542 conventional metaphors ($M = 2.97, SE = .08$), ($p < .001$), with literal expressions being most familiar
 543 ($M = 3.85, SE = .09$), ($p < .001$). Furthermore, conventional metaphors were less familiar than literal
 544 word dyads, ($p < .001$). It should be noted that all word pairs were relatively unfamiliar, which is evident

Table 3 Accuracy scores (for the class targets) and Pearson correlations (for the graded ratings) for semantic features of verb-noun pairs.

	class	metaphoricity	meaningfulness	familiarity
all others	0.737	0.686	0.734	0.714
metaphoricity	0.715	-	-0.641	-0.613
meaningfulness	0.579	-0.641	-	0.675
familiarity	0.583	-0.613	0.675	-

545 in the mean score for literal word pairs. They were evaluated as most familiar of all three categories,
 546 but did not obtain maximum familiarity values on the scale (below 4, while 6 and 7 represented highly
 547 familiar items). Familiarity was low in all three categories as we intentionally excluded highly probable
 548 combinations.

549 4.1.4 Metaphoricity

550 In order to assess the metaphoricity of the word pairs, participants decided how metaphoric a given
 551 word dyad was on a scale from 1 (very literal) to 7 (very metaphoric). A main effect of utterance type
 552 was found, [$F(2, 198) = 588.82, p < .001, \epsilon = .738, \eta_p^2 = .86$]. Pairwise comparisons showed that novel
 553 metaphors ($M = 5.00, SE = .06$) were rated as more metaphoric than conventional metaphors ($M = 3.98,$
 554 $SE = .06$), ($p < .001$), and conventional metaphors were rated as more metaphoric than literal utterances
 555 ($M = 2.74, SE = .07$), ($p < .001$).

556 4.2 Correlations in Human Judgements

557 In order to understand the way in which meaningfulness, familiarity, and metaphoricity interact in the
 558 judgements reported by humans, we model the correlations between each of these factors, as well as the
 559 propensity of each of these factors to identify the metaphoric class of a phrase (that is, whether it is literal,
 560 conventional, or novel). Results are reported in Table 3.

561 The accuracy ratings for class are determined by performing a logistic regression taking the graduated
 562 human ratings for each semantic category as independent variables. Membership of each of the three
 563 candidate classes is determined through a one-versus-rest scheme; the results in the class column of
 564 Table 3 are based on a leave-one-out cross-validation. In the case of *all others*, each of the three different
 565 semantic categories serve as the independent variables in a multi-variable logistic regression. Unsurprisingly,
 566 metaphoricity itself is most predictive of the metaphoric class of a phrase ($p = .054$ for the difference
 567 between metaphoricity and familiarity, based on a permutation test). The enhancement in accuracy by
 568 adding familiarity and meaningfulness to the model based only on metaphoricity is, on the other hand, not
 569 significant ($p = .574$).

570 Figure 3 seeks to visualise the relationship between metaphoricity and the other two semantic phenomena
 571 measured here by projecting metaphoric classes of verb-object phrases in terms of meaningfulness and
 572 familiarity. The correlation between increases in familiarity and meaningfulness and the drift from literal
 573 phrases through conventional metaphors to novel metaphors is apparent, though there is also a good deal of
 574 overlap in the scores assigned to each category, with outliers from each class to found at all extents of the
 575 statistical cluster.

576 There are plenty of phrases that are considered meaningful but unfamiliar, and these phrases tend to be
 577 considered either literal or conventionally metaphoric, but there are very few phrases that are considered
 578 familiar and meaningless. It is tempting to therefore hypothesise that we might construe familiarity as, in

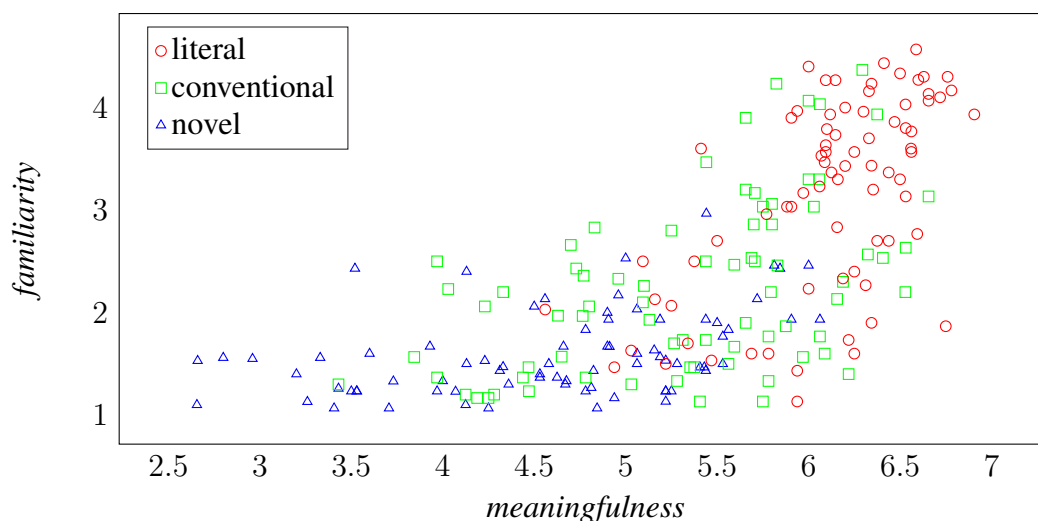


Figure 3 The three metaphoric classes as functions of meaningfulness and familiarity.

579 itself, a product of meaning: there is an inherent relationship by which recognising a semantic composition
 580 is contingent on recognising its meaningfulness. More pertinently, we will claim that the process by which
 581 metaphor emerges from a cognitive re-representation of the world is evident in the way that humans judge
 582 these assessments of semantic categories to play out across these three classes of verb-object phrases.
 583 Those phrases that veer into the unfamiliar in particular are associated with the conceptual contortions
 584 implicit in novel metaphor.

5 EXPERIMENTAL METHODOLOGY

585 Building on the methodology for constructing a base space, projecting contextually informed subspaces
 586 from this base space, and extracting geometric features suitable for semantic analysis from these subspaces,
 587 we now turn to the project of applying this methodology to a model that captures the semantic assessments
 588 of humans. We apply the techniques outlined in Section 3 to generate geometries associated with input
 589 in the form of verb-object phrases. We are effectively testing the degree to which human judgements of
 590 metaphor can be captured in statistical observations of word co-occurrences, and then exploring how these
 591 statistical tendencies can be contextually projected onto geometric features. Our modelling methodology
 592 will involve learning linear mappings between geometric features and human scores, as well as logistic
 593 regressions designed to predict metaphoric class.

594 In practice, this involves producing subspaces associated with each of the verb-object dyads in the dataset
 595 described in Section 4. In these subspaces, the words composing the dyad are represented as vectors,
 596 and these vectors have a geometrical relationship to one another and to the subspace itself which can be
 597 represented as a feature vector (corresponding to the features described in Table 1). Our hypothesis is that
 598 these geometric features, which are designed to represent the semantics of the particular context associated
 599 with each input dyad, will map to ratings regarding the metaphoricity, meaningfulness, and familiarity of
 600 the dyad in question. This, returning to the theoretical background of Section 2.3 and model of Section 3.1,
 601 is intended to provide a computational mechanism that is conducive to modelling metaphor as a process of
 602 *ad hoc* concept construction within a particular communicative context.²

² Scripts for building dynamically contextual distributional semantic models, as well as for using these models to project context-specific subspaces and use these subspaces to model human metaphor judgements, are available at <https://github.com/masteradamo/metaphor-geometry>. The data on human

603 5.1 Modelling metaphoric re-representation from geometries of subspaces

604 We begin our experiments by building a base space of word-vectors based on a statistical analysis of
 605 Wikipedia, as described in Section 3.2: this results in a matrix of information theoretical co-occurrence
 606 statistics. This matrix will serve as the basis for projections contextualised by particular verb-object
 607 compositions. In order to model the relationship between lexical semantic representations re-represented in
 608 potentially metaphoric contexts, we take each word pair in the dataset described in Section 4.1 as input to
 609 each of the three subspace projection techniques described in Section 3.3, working off the base space to
 610 generate 200 dimensional subspaces. We project the word-vectors associated with each input word into
 611 each subspace, and also compute the mean-vector, maximum-vector, and central-vector for each subspace.
 612 Based on these projections, we calculate the 48 geometric features listed in Table 1.

613 These features are then used as independent variables in least squares regressions targeting the human
 614 ratings for each of the three semantic categories assessed for each verb-object phrase: metaphoricity,
 615 meaningfulness, and familiarity.³ We pre-process the geometric measures by performing mean-zero,
 616 standard-deviation-one normalisation across each feature. We similarly perform a logistic regression on the
 617 same normalised matrix of geometric features to learn to predict the metaphoric class (literal, conventional,
 618 or novel) of each dyad in our data. As with the model mapping from semantic ratings to classes described
 619 in Section 4.2, we employ a one-versus-rest scheme, so in effect we fit three different models, one for each
 620 class, and then classify a phrase based on the model for which that phrase scores highest.⁴ We once again
 621 employ a leave-one-out cross-validation technique.

622 The objective here is to evaluate the extent to which the geometric features of the subspaces we project
 623 collectively capture the contextual semantics of a particular dyad. By evaluating each dyad d on a regression
 624 of the the 227×48 matrix of independent variables D' , defined such that $d \notin D'$ (227 for all the dyads
 625 in our datasete except d , and 48 for the entire set of geometric features defined in Table 1), and then
 626 aggregating the average correlation scores across all dyads, we can get a general picture of the degree to
 627 which these features collectively correlate with human judgements.

628 5.2 Semantic Geometry

629 The full-featured approach described above offers a good overall sense of the way that statistical geometry
 630 maps to semantic features. There will, however, be a good deal of collinearity at play in the geometric
 631 features we have defined for our model. The angle between the verb and noun vectors, for instance ($\angle VON$
 632 in Figure 2) would be expected to correlate somewhat with \overline{VN} , the Euclidean distance between the vectors.
 633 Likewise, the ratio of the smaller to the larger of distances between the word-vectors and the mean-vector
 634 $\overline{VM} : \overline{NM}$ will in many subspaces be identical to the fraction $\overline{VM}/\overline{NM}$.

635 To address this, we undertake a feature-by-feature analysis of our data. We isolate each of the 48 geometric
 636 features listed in Table 1 and calculate the Pearson correlation between the feature and the human ratings
 637 for each of the three semantic phenomena under consideration. This move provides the basis for an analysis
 638 of the way that specific aspects of the geometry of a contextualised subspace map to human judgements,
 639 which in turn allows us to tease out the specific correlations between co-occurrence statistics observed in a
 640 large-scale corpus and the re-representational processes associated with metaphor interpretation. In this

metaphor judgements is available at https://figshare.com/articles/To_Electrify_Bilingualism_Electrophysiological_Insights_into_Bilingual_Metaphor_Comprehension/4593310/1; this data is described in detail by Jankowiak et al. (2017).

³ This is implemented using the sklearn `LinearRegression` module for Python.

⁴ This implements using the sklearn `LogisticRegression` module for Python.

641 sense, our subspace architecture becomes a geometric index mapping from the unstructured data available in
642 a corpus to the dynamics of language in use.

643 5.3 Eliminating Collinearity

644 As mentioned above, there is inevitably collinearity between the geometric features we use to give
645 analytical structure to our subspaces. Among other things, features corresponding to points of the
646 normalised component of the geometry (so, V' , C' , M' , X' , and C') will in many cases correlate
647 with corresponding features associated with the non-normalised component of the geometry. In order to
648 overcome this aspect of our geometric data, we apply a variance inflation factor to construct a reduced set
649 of truly independent variables (O'Brien, 2007). This is effectively a statistic computed to iteratively build
650 up a vector of adequately non-correlated geometric features by assessing the degree of covariance each
651 additional feature would introduce to the aggregating set of features.

652 Our process begins by seeding an input matrix with the measures for each verb-object phrase for the
653 top ranking geometric feature for a given semantic phenomena. We then move down the list of features,
654 calculating the coefficient of determination R^2 for a least squares linear regression between the established
655 matrix and the measures associated with the next variable. We concatenate the next variable to our list of
656 independent variables only if the following criterion is met:

$$\frac{1}{1 - R^2} < fac \quad (2)$$

657 We set the model parameter fac at the quite stringent level of 2, and then select up to 5 out of the 48
658 features outlined in Table 1 as the independent variables for a linear regression trained on human ratings
659 for three different semantic categories. We use this non-collinear set of features to run linear and logistic
660 regressions to learn to predict semantic phenomena and metaphoric class respectively, applying once again
661 leave-one-out cross-validations. This process results in a set of geometric features that we expect to be
662 optimally informative in terms of correlations with human semantic judgements. This should offer us an
663 opportunity to analyse in more detail the interactions between different features.

6 RESULTS

664 Having established our experimental methodology, we apply the three different empirical stages outlined
665 in Section 5: a full-featured cross-evaluation of linear models mapping from the geometries of subspaces
666 to human judgements of metaphoricity, meaningfulness, and familiarity; cross-evaluations of feature-by-
667 feature linear models; and finally cross-evaluation of linear models constructed based on an iterative
668 analysis designed to minimise collinearity between selected geometric features. Here we present results,
669 with statistical significance calculated where appropriate, in terms of Fisher r-to-z transforms for rating
670 correlations and permutation tests for classification f-scores.

671 6.1 Multi-Feature Correlations

672 Results for experiments involving linear models mapping all 48 geometric features of subspaces to graded
673 human judgements of metaphoricity, meaningfulness, and familiarity are reported in the first three rows of
674 Table 4. In the last row, labeled "class", accuracy results for a logistic regression mapping from the full set
675 of geometric features to human classifications of verb-object dyads as literal non-metaphors, conventional
676 metaphors, or novel metaphors are reported. For these multi-feature correlations, we report results for

Table 4 Pearson correlations for leave-one-out cross-validated linear regressions predicting semantic judgements based on geometric features extrapolated using three different subspace selection techniques, as well as with cosine similarity for the WORD2VEC baseline. This is followed by accuracy for predicting the correct metaphoric class for each phrase.

	INDY	MEAN	GEOM	W2V	single-class baseline
metaphoricity (correlation)	0.442	0.348	0.419	-0.288	-
meaningfulness (correlation)	0.430	0.380	0.290	0.215	-
familiarity (correlation)	0.452	0.283	0.391	0.224	-
class (accuracy)	0.447	0.447	0.442	0.458	0.333

677 all three subspace projection techniques: subspaces delineated by co-occurrence features independently
 678 selected based on the profile of each word in a dyad, and then subspaces selected based on the arithmetic
 679 and geometric means of co-occurrence features between the input words in a dyad.

680 Interestingly, the features generated by the INDY technique most closely reflect human judgements for all
 681 three semantic categories (though, even for the largest difference between the INDY and MEAN techniques
 682 for familiarity, significance is marginal at $p = .038$ for a Fisher r -to- z transform). This is a bit less evident
 683 in terms of metaphoricity, where the GEOM technique achieves an appreciable correlation; nonetheless, it
 684 would appear that subspaces generated from the conjunction of dimensions independently salient to each
 685 of the two words involved in a phrase provide the most reliable geometric basis for predicting how humans
 686 will judge the phrase.

687 The results for predicting class are not significantly above the baseline accuracy score of 0.333 (indicated
 688 in the fifth column of Table 4), which would entail, for instance, predicting every phrase to be literal ($p =$
 689 $.092$ for the difference between this baseline and the INDY output, based on a permutation test). Beyond
 690 that, the different subspace selection techniques are more or less in line with one another, suggesting that,
 691 more than for graduated human ratings of semantic phenomena, there is not much to choose between the
 692 different geometries generated here—at least when they are taken as a relatively high dimensional set of
 693 features entered into a regression model.

694 We compare these results with correlations and a logistic regression derived from the `word2vec` model
 695 described in Section 3.5. As cosine similarity is the singular measure for judging the relationship between
 696 two words, we simply calculate the Pearson correlation between pairs of words in our input phrases and
 697 human ratings for the three graded semantic phenomena. We likewise perform a one-versus-rest multi-class
 698 logistic regression to learn to predict the metaphoric class for each phrase. Results are reported in the fourth
 699 column of Table 4. The difference in metaphoricity scores between correlations with the INDY technique
 700 and the `word2vec` baseline are not significant ($p = .059$ based on a Fisher r -to- z transform). Furthermore,
 701 `word2vec` is actually better at predicting the metaphoric class of a phrase than the model trained on all
 702 the geometric features of our model.

703 6.2 Single-Feature Correlations

704 There are a very large number of single-feature correlations to analyse: 48 separate ones, one for each
 705 component of the geometric feature map illustrated in Figure 2 and detailed in Table 1, multiplied by three
 706 different subspace projection techniques. We focus on the features extracted from subspaces generated
 707 using the INDY technique, as the initial results from Table 4 suggest that these subspaces might be the most
 708 interesting from a semantic perspective. The top five features, in terms of the absolute value of correlation,
 709 are reported in Table 5, using the geometric nomenclature from Table 1 with reference to Figure 2.

Table 5 Top independent geometric features for three semantic phenomena as found in INDY subspaces, ranked by absolute value of Pearson correlation.

metaphoricity		meaningfulness		familiarity	
$\angle VON$	-0.524	$\angle VON$	0.451	$\angle VMN$	0.431
$\overline{V'N'}$	0.519	$\overline{V'N'}$	-0.447	$\angle VCN$	0.425
$\mu(\overline{V'C'}; \overline{N'C'})$	0.509	$\mu(\overline{V'M'}; \overline{N'M'})$	-0.437	$\mu(\overline{VC}; \overline{NC})$	-0.418
$\mu(\overline{V'M'}; \overline{N'M'})$	0.506	$\Delta V X N$	-0.435	$\overline{V'N'}$	-0.407
$\Delta V X N$	0.504	$\mu(\overline{V'C'}; \overline{N'C'})$	-0.433	$\angle VON$	0.406

710 Not surprisingly, there is a degree of symmetry here: the results for metaphoricity and meaningfulness
 711 in particular come close to mirroring one another, with strongly positive correlations for one phenomena
 712 being strongly negative for the other, in line with the negative correlations between these phenomena as
 713 reported by humans in Table 3. The angle between the word-vectors, for instance ($\angle VON$), correlates
 714 negatively with metaphoricity and positively with meaningfulness. This makes sense when we consider
 715 that a cosine relatively close to 1 between two vectors means that they are converging in a region of
 716 a subspace (regardless of their distance from the vector), and aligns with the strong results for cosine
 717 similarity achieved by our `word2vec` model, accentuated by the contextualisation afforded by the INDY
 718 contextualisation technique.

719 What is perhaps surprising about these results is that there is such a clear, albeit inverse, correlation
 720 between the features that indicate metaphoricity and meaningfulness in these subspaces, while familiarity
 721 is associated with a slightly different geometric profile. This finding in regard to familiarity seems to
 722 stem from the non-normalised region of the subspace, suggesting that word-vectors that are not only
 723 oriented similarly but also have a similar relationship to the origin are more likely to be considered
 724 familiar. It would seem, then, that, in terms of the relationships between metaphoricity and meaningfulness,
 725 directions in a subspace are indicative of the semantic shift from the meaningful and known to metaphoric
 726 re-representation.

727 6.3 Optimised Correlations

728 Moving on from the single-feature analysis of each geometric feature of a particular type of subspace
 729 projection, we now turn to models built using multiple independent geometric features selected based on
 730 their independent performance constrained by a variance inflation factor, as described in Section 5.3. To
 731 recapitulate, this involves adding one-by-one the top features as returned by the single-feature analysis
 732 reported above, so long as each additional feature does not exceed a value of 2 for the measure *fac*
 733 formulated in Equation 2, until at most five features are included in the optimised space of geometric
 734 features. Overall results for each subspace projection technique are reported in Table 6.

735 Once again, the INDY projection technique outperforms the other two techniques, as well as the the
 736 `word2vec` baseline on all counts, including now accuracy of classification of verb-object dyads. There is
 737 a marked improvement for both the INDY and MEAN techniques ($p = .080$ for the difference between the
 738 non-optimised and optimised INDY metaphoricity predictions). The INDY results are also improvements
 739 on the best scores for individual geometric features reported in Table 5, though the difference here is less
 740 pronounced. But on the whole, for these two techniques, there is clearly some advantage to discovering a
 741 set of non-collinear geometric features in order to understand how distributional statistics can be mapped
 742 to semantic judgements. Moreover, this refined version of our model outperforms the `word2vec` baseline

Table 6 Pearson correlations for leave-one-out cross-validated linear regressions predicting human judgements based on geometric features extrapolated using three different subspace selection techniques with up to 5 independent geometric features selected based on a variance inflation factor.

	INDY	MEAN	GEOM	W2V	single-class
metaphoricity (correlation)	0.565	0.447	0.305	-0.288	-
meaningfulness (correlation)	0.492	0.428	0.255	0.215	-
familiarity (correlation)	0.464	0.383	0.318	0.224	-
class (accuracy)	0.531	0.465	0.412	0.458	0.333

743 in all regards, including prediction of metaphoric class, though the difference is not statistically significant
744 ($p = .247$ for the difference between the INDY technique and `word2vec`).

745 It is nonetheless interesting that a reduction in features motivated by observations about particular aspects
746 of semantic geometry actually gives us a more productive model. As Guyon and Elisseff (2003) point out,
747 this is possibly an indicator of an underlying non-linearity between the geometric features of our subspaces
748 and the human judgement of semantic properties. Given this, we may expect further improvement in results
749 using for instance a neural modelling technique, but here our intentions are to explore the geometry of the
750 subspaces in a straightforward and interpretable way, so we leave explorations of more computationally
751 complex modelling for future study.

752 Table 7 focuses on the top features for each phenomenon as selected for the INDY technique in particular.
753 There are some telling trends here: where distance $\overline{V'N'}$ was independently predicative of all three semantic
754 criteria in Table 5, this is hedged out by the even more predictive cosine measure $\angle VON$ for metaphoricity
755 and meaningfulness, because the correlation between $\overline{V'N'}$ and $\angle VON$ is too high to satisfy *fac*. That
756 these measures both correlate positively with meaningfulness is telling us that word-vectors detected to the
757 same side of the middle of a subspace are more likely to form a meaningful composition and less likely to
758 form a metaphorical one, but the presence of both of them in our analysis doesn't tell us much that the
759 presence of one or the other wouldn't. A similar story can be told for the positive correlation of the angles
760 at the vertices of both non-normalised mean and central vectors in the case of familiarity ($\angle VMN$ versus
761 $\angle VCN$). Again, it's not particularly surprising to see features like the mean distance between normalised
762 word vectors and both normalised mean and central vectors achieving similar scores ($\mu(\overline{V'M'}; \overline{N'M'})$
763 versus $\mu(\overline{V'C'}; \overline{N'C'})$).

764 To assess this final step in our modelling process in a little more detail, we consider the features themselves,
765 along with the coefficients assigned to them in an all-in linear regression. These values are listed for the
766 INDY technique in Table 7. We once again note a strong negative correlation between the features that
767 select for metaphoricity versus the features that select for meaningfulness, with word-vectors that are found
768 at wide angles (based on the $\angle VON$ feature) and at relatively different distances from generic vectors
769 (based on the $\overline{VX}/\overline{NX}$ and $\overline{VX} : \overline{NX}$ features) more likely to form a metaphoric composition.

770 Familiarity indicates a somewhat similar profile of features: like with meaningfulness, subspaces where
771 the verb-vector and noun-vector are, on average, closer to the maximum extent of the space (X) tend
772 to indicate a composition which humans will consider more familiar. The positive correlation of the
773 fraction $\overline{VC}/\overline{NC}$ actually makes sense in relation to the (marginally) negative correlation with the fraction
774 $\overline{VX}/\overline{NX}$, because we can expect to generally find the word-vectors that select these subspaces in the region
775 between the central-vector C and the maximum-vector X . So it would seem that, as with meaningfulness,
776 as the verb-vector grows relatively closer to X compared to the noun-vector, phrases are more likely to be
777 familiar to humans.

Table 7 Top geometric features for three semantic phenomena as found in INDY subspaces, ranked in the order that they are selected based on a variance inflation factor criterion, along with coefficients assigned in an all-in linear regression.

metaphoricity		meaningfulness		familiarity	
$\angle VON$	-0.297	$\angle VON$	0.134	$\angle VMN$	0.296
$\mu(\overline{VX}; \overline{NX})$	0.067	$\mu(\overline{VX}; \overline{NX})$	-0.111	$\mu(\overline{VX}; \overline{NX})$	-0.168
$\angle V'X'N'$	-0.150	$\angle V'X'N'$	0.157	ΔVMN	0.005
$\overline{VX}/\overline{NX}$	0.217	$\overline{VX}/\overline{NX}$	-0.249	$\overline{VC}/\overline{NC}$	0.184
$\overline{VX} : \overline{NX}$	0.162	$\overline{V'C'} : \overline{N'C'}$	-0.205	$\overline{VX}/\overline{NX}$	-0.050

7 DISCUSSION

778 Having established the results of our dynamically contextual methodology's ability to model human
 779 judgements of metaphoricity, meaningfulness, and familiarity, we turn to an analysis of the components of
 780 our experimental set-up. In addition to an overall assessment of the methodology and a consideration of
 781 performance of certain parameter settings and particular geometric features, we would like to emphasise
 782 the way that the combination of subspace projection and linear feature mapping works to provide the
 783 framework for a more nuanced consideration of the relationship between corpus analysis and the cognitive
 784 and linguistic components of semantic phenomena. Our overall claim is that the context-specific and
 785 geometrically nuanced approach we have endorsed here shows promise as a way for using computational
 786 modelling to explore language as a fundamental component of human behaviour.

7.1 Model Parameters

788 One of the findings that emerges from the results presented in Section 6 is an opportunity to compare
 789 different modelling parameters, and to consider the relationship between these components of our
 790 methodology and metaphoric re-representation. The modelling feature that is of most interest here is
 791 the difference between the INDY, MEAN, and GEOM subspace projection techniques, and the primary thing
 792 to note is the superior performance of the INDY technique in modelling human considerations of all three
 793 semantic phenomena investigated here: metaphoricity, meaningfulness, and familiarity.

794 We begin by recalling that, as mentioned in Section 3.3, the MEAN and GEOM techniques are really two
 795 different ways of computing average values of co-occurrence features potentially shared between different
 796 input words, while the INDY technique produces a subspace that is a mixture of co-occurrence features
 797 that are independently salient to one word or the other—or possibly, but not necessarily, both. In fact,
 798 what we might be seeing in the strong correlations between geometric features of the INDY subspaces and
 799 human judgements is, in part, the identification of instances where the co-occurrence profiles of input
 800 words tend to converge or diverge. This claim is supported by the strong negative correlation between
 801 metaphoricity and cosine ($\angle VON$) in Table 7, along with the positive correlation with the mean distance of
 802 the vectors from the maximal point X , and the opposite set of correlations for the same features observed
 803 for meaningfulness. As the set of independently selected co-occurrence features evidence less overlap for
 804 the two components of the verb-object input dyad, the angle of the contextually projected word-vectors
 805 corresponding to these inputs drift apart in the subspace, and the regions of the projection become less
 806 correspondent with one another.

807 Additionally, the GEOM methodology actually realises lower Pearson correlations for non-collinear
 808 combinations of geometric features than it does for the full set of geometric features. The definitive aspects
 809 of this technique are that it only selects co-occurrence dimensions with non-zero values for both input

810 words, and that it furthermore tends to favour dimensions where the value is pretty high for both input
811 words rather than very high for one and not so high for the other (the geometric mean of (5,5) is 5, but for
812 (9,1) it is only 3). These subspaces therefore should already exhibit a good degree of information about
813 both word-vectors of a verb-object phrase, so there is perhaps less to be discovered in measures such
814 as angular divergences relative to generic vectors near the centre of a subspace. On the other hand, the
815 requirement for mutually non-zero co-occurrence dimensions means that co-occurrences with relatively
816 common words will eventually have to be selected, and so we might find information about co-occurrence
817 features that are not in any sense conceptually salient, but instead just happen to come up quite often in
818 our corpus. We could hypothesise that a larger co-occurrence window would yield stronger predictions for
819 these subspaces, since there would be more observations of co-occurrences in the corpus for any given
820 word-vector. We leave further experimentation along these lines for future work.

821 **7.2 Using Geometry to Interpret Semantics**

822 The analysis offered above of the strong performance of the INDY subspace selection technique is
823 indicative of the general way in which we would like to suggest that statistical geometries can be
824 mapped to semantic phenomena. The combination of interpretable projections and nuanced analysis
825 of the way that input word-vectors tend to move around relative to contexts associated with a set of graded
826 semantic measures turns the list of geometric features enumerated in Table 1 into a set of semantic indices,
827 providing traction for using modelling techniques that move from statistics about word co-occurrences to
828 commitments about the way that humans use metaphor. In this way, geometric analysis maps to cognitive
829 phenomena, elevating the model from something that merely learns to predict correlations to something that
830 captures the way concepts are manipulated and indeed generated in response to an unfolding environment.

831 The divergence between the relatively congruent, albeit converse, features that model metaphoricity and
832 meaningfulness as compared to the features that model familiarity offers a case in point. There is a close
833 semantic relationship between metaphor and meaning: we might argue that a metaphor involves shifting a
834 concept to suit a situation, and new meaning is produced as a result of this shifting. Familiarity, on the
835 other hand, is an epistemological phenomenon with a frequentist connotation, and so is not expected to
836 map neatly to this relationship between metaphor and meaning. This disconnect seems to play out in the
837 interpretable geometry of context specific subspaces projected by our model. In the geometric features that
838 provide traction to our model, the non-linear tension between familiarity and meaningfulness as reported by
839 humans and illustrated in Figure 3 is teased out in terms of the distinct set of geometric features associated
840 with familiarity. In particular, in Tables 5 and 7, we see that familiarity has a relationship with the mean
841 point M in contextual subspaces, suggesting that the relationship between projected word-vectors relative
842 to the typical non-zero characteristics of a projection tell us something about how readily accepted a
843 composition will be to humans.

844 **7.3 The Dynamic Geometry of Representation**

845 In order to examine more closely the nature of re-representation by way of contextualised projections of
846 statistical geometry, we look at two case studies. Each case involves one noun applied to three different
847 verb-object phrases, one judged to be literal, one conventionally metaphoric, and one a novel metaphor, as
848 outlined in Section 4.1. Our objective is to offer a qualitative, visually grounded analysis of the way that
849 the typical geometry of projections shifts as we move across the spectrum of metaphoricity.

850 Our two examples are presented in Figure 4, where the word-vectors and generic vectors as projected
851 into 200 dimensional subspaces using the INDY subspace selection technique are further projected into

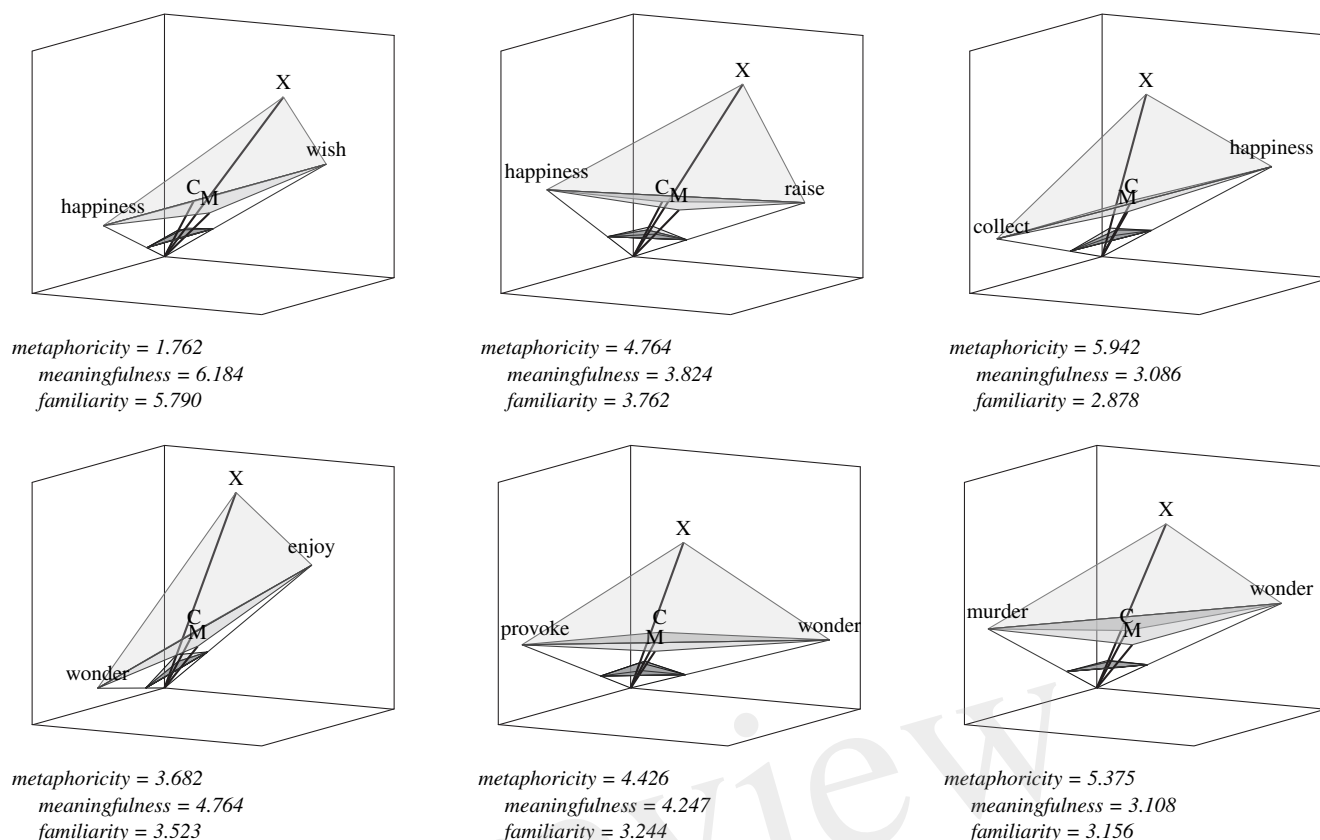


Figure 4 Subspaces, including word-vectors and generic features, for two different nouns composed with three verbs each, ranging from literal on the left to novel metaphor on the right. These three-dimensional projections have been derived through a regression designed to preserve the norms of all vectors, the distances between the word vectors, and the distances between each word-vector and all the generic vectors. The ratings assigned by our model are indicated below each plot.

852 perspectives on three-dimensional renderings. These instances have been selected because the ratings
 853 output for metaphoricity by our model follow a regular progression as we move from literal to conventional
 854 to novel compositions. The first example involves the phrases *wish happiness*, *raise happiness*, and
 855 *collect happiness*; the second example involves the phrases *enjoy wonder*, *provoke wonder*, and
 856 *murder wonder*. With each noun, metaphoricity as rated by our model progressively increases with each successive
 857 composition, and meaningfulness and familiarity conversely decrease.

858 Along with this progression, we observe a gradual expansion of the complexes of vectors as we move from
 859 the literal to the overtly metaphoric. This is in line with the widening of the angle $\angle VON$, as statistically
 860 observed in Table 6. We also note an extension of the maximal-vector X away from the other points of
 861 interest in a subspace, a characteristic predicted by the increase of the mean distance between the word
 862 vectors and the maximal-vector $\mu(\overline{VX} : \overline{NX})$. In terms of the spreading of the angle $\angle VMC$ characteristic
 863 of decreasing familiarity, this is harder to perceive in this visualisation, but there is a detectable flattening
 864 of the already wide vertices at both M and C by the time we get to *collect happiness* in particular.

865 In the end, it is difficult to make any very precise observations about these figures. They are necessarily
 866 lossy projections from much higher dimensional spaces, and the tricks of perspective when rendering
 867 three dimensions onto a plane also means that information about angular relationships even in these
 868 low-dimensional projections is easily lost. The purpose of these last illustrations is not so much to provide
 869 a tool for rigorous quantitative analysis, which has been provided above, as to show in a more general and

870 qualitative sense that there is a spatial quality to the way that metaphor emerges as we edge away from the
871 familiar and the meaningful. We argue that this quality corresponds to the re-representation inherent in
872 constructing novel ways of talking about situations in the world.

873 Perhaps the appropriate way to think about metaphoric re-representation is in terms of a discovery of
874 unfamiliar meaning in a particular context. So, while both humans and our computational model tend
875 to identify a negative correlation between meaningfulness and metaphoricity, we could imagine how
876 phrases like *collect happiness* and *murder wonder* could gain potent semantics in the right situation.
877 Our computational model, underwritten by concrete and quantifiable observations of the way that words
878 tend to be used, is designed to extrapolate a more general geometric way of capturing the process
879 by which contextualisation leads to the *ad hoc* construction of new representations with very specific
880 communicative potentialities. Without wanting to make too strong a claim about what we can expect
881 from computational models, we suggest that this geometric mode of representing metaphor in terms of
882 statistical information about large-scale co-occurrence tendencies hints at a move towards a computational
883 methodology for capturing some of the non-propositional and phenomenological components of figurative
884 language (Davidson, 1978; Reimer, 2001; Carston, 2010b).

8 CONCLUSION

885 We argue here that dynamically projecting context-specific conceptual subspaces into new representations
886 captures the mapping process that is necessary for conceptually resolving the semantics of non-literal
887 language. We hypothesised that the geometry defining these subspaces (which reflects lexical co-occurrence
888 relationships in a large-scale textual corpus) can be thought of as a quantification of the process of
889 re-representation. This allows us to examine how the conceptual re-mappings underlying metaphoric
890 language perception are related to underlying mathematically-tractable lexical semantic representations.
891 By examining features of contextualised subspaces, our novel methodology can be used to assess the way
892 that the overall geometric quality of a representation in our model maps to metaphoric shifts in meaning.
893 We believe that this aspect of our approach may point the way towards the computational modelling of
894 some of the more elusive theoretical properties of figurative language as a cognitive mechanism for moving
895 away from propositional content.

896 Our methodology has been designed to accommodate pragmatic accounts of metaphor, by which figurative
897 compositions involve the construction of an *ad hoc* conceptual space: the subspaces projected by our
898 dynamically contextual model correspond to these extemporaneously projected semantic relationships. This
899 facility is not intended to come at the expense of other accounts of metaphor; rather, we have been motivated
900 by exploring ways that a theoretical stance that has typically proved challenging for computational semantic
901 modelling can be addressed within the broader paradigm of distributional semantics.

902 With this in mind, we can imagine ways that future development of our methodology might lend itself to
903 practical applications in neurolinguistic and clinical contexts. For instance, experimental evidence indicates
904 major deficits in metaphoric language in conditions such as schizophrenia (Bambini et al., 2016): our
905 methodology could provide a quantitative tool for introducing this pragmatic component to predict clinical
906 diagnosis, as proposed for other aspects of language (Foltz et al., 2016). More generally, our approach can
907 be counted as a contribution to a growing body of literature that seeks to use data-drive techniques to make
908 links between neurolinguistic studies and some of the more complex aspects of language in use (Jacobs
909 and Kinder, 2017), epitomised by the contextually situated re-representation at play in the use of metaphor.

ACKNOWLEDGEMENTS

910 This research was partially supported by the project ConCreTe, which acknowledges the
 911 financial support of the Future and Emerging Technologies (FET) programme within
 912 the Seventh Framework Programme for Research of the European Commission under
 913 FET grant number 611733; and by the European Union's Horizon 2020 research and
 914 innovation programme under grant agreement No 825153, project EMBEDDIA (Cross-
 915 Lingual Embeddings for Less-Represented Languages in European News Media). The
 916 results of this publication reflect only the authors' views and the Commission is not responsible for any use
 917 that may be made of the information it contains. This research has also been supported by EPSRC grant
 918 EP/L50483X/1 and by the CHIST-ERA project ATLANTIS.



CONTRIBUTION

919 All authors have contributed to both the writing of this paper and the underlying research.

REFERENCES

- 920 Agres, K., McGregor, S., Purver, M., and Wiggins, G. (2015). Conceptualising creativity: From
 921 distributional semantics to conceptual spaces. In *Proceedings of the 6th International Conference*
 922 *on Computational Creativity* (Park City, UT), 118–125
- 923 Agres, K. R., McGregor, S., Rataj, K., Purver, M., and Wiggins, G. (2016). Modeling metaphor
 924 perception with distributional semantics vector space models. In *Proceedings of the ESSLLI Workshop*
 925 *on Computational Creativity, Concept Invention, and General Intelligence (C3GI)*, eds. T. R. Besold,
 926 O. Kutz, and C. Leon (Bolzano-Bozen, Italy), 1–14
- 927 Arzouan, Y., Goldstein, A., and Faust, M. (2007). Brainwaves are stethoscopes: ERP correlates of novel
 928 metaphor comprehension. *Brain Research* 1160, 69–81
- 929 Bambini, V., Arcara, G., Bechi, M., Buonocore, M., Cavallaro, R., and Bosia, M. (2016). The
 930 communicative impairment as a core feature of schizophrenia: Frequency of pragmatic deficit,
 931 cognitive substrates, and relation with quality of life. *Comprehensive Psychiatry* 71, 106–120.
 932 doi:10.1016/j.comppsy.2016.08.012
- 933 Bambini, V., Canal, P., Resta, D., and Grimaldi, M. (2019). Time course and neurophysiological
 934 underpinnings of metaphor in literary context. *Discourse Processes* 56, 77–97
- 935 Barnden, J. A. (2008). Metaphor and artificial intelligence: Why they matter to each other. In *The*
 936 *Cambridge handbook of metaphor and thought*, ed. J. E. R. W. Gibbs (Cambridge University Press).
 937 311–338
- 938 Barsalou, L. W. (1999). Perceptions of perceptual symbols. *Behavioral and Brain Sciences* 22, 637–660
- 939 Bowdle, B. F. and Gentner, D. (2005). The career of metaphor. *Psychological Review* 112, 193
- 940 Brouwer, H., Crocker, M. W., Venhuizen, N. J., and Hoeks, J. C. (2017). A neurocomputational model of
 941 the n400 and the p600 in language processing. *Cognitive science* 41, 1318–1352
- 942 Brouwer, H. and Hoeks, J. C. (2013). A time and place for language comprehension: mapping the n400
 943 and the p600 to a minimal cortical network. *Frontiers in human neuroscience* 7, 758
- 944 Carston, R. (2010a). Lexical pragmatics, ad hoc concepts and metaphor: A relevance theory perspective.
 945 *Italian Journal of Linguistics* 21, 153–180
- 946 Carston, R. (2010b). Metaphor: Ad hoc concepts, literal meaning and mental images. *Proceedings of the*
 947 *Aristotelian Society* CX, 297–323

- 948 Clark, S. (2015). Vector space models of lexical meaning. In *The Handbook of Contemporary Semantic*
949 *Theory*, eds. S. Lappin and C. Fox (Wiley-Blackwell). 2nd edn., 493–522
- 950 Coecke, B., Sadzadeh, M., and Clark, S. (2011). Mathematical foundations for a compositional distributed
951 model of meaning. *Linguistic Analysis* 36, 345–384
- 952 Coulson, S. and Van Petten, C. (2002). Conceptual integration and metaphor: An event-related potential
953 study. *Memory & cognition* 30, 958–968
- 954 Davidson, D. (1978). What metaphors mean. *Critical Inquiry* 5, 31–47
- 955 De Grauwe, S., Swain, A., Holcomb, P. J., Ditman, T., and Kuperberg, G. R. (2010). Electrophysiological
956 insights into the processing of nominal metaphors. *Neuropsychologia* 48, 1965–1984
- 957 Deerwester, S., Dumais, S. T., Furnas, G. W., Landauer, T. K., and Harshman, R. (1990). Indexing by
958 latent semantic analysis. *Journal for the American Society for Information Science* 41, 391–407
- 959 Foltz, P. W., Rosenstein, M., and Elvevåg, B. (2016). Detecting clinically significant events through
960 automated language analysis: Quo imus? *Npj Schizophrenia* 2, 15054 EP –
- 961 Gibbs, R. W. and Tendahl, M. (2006). Cognitive effort and effects in metaphor comprehension: Relevance
962 theory and psycholinguistics. *Mind & Language* 21, 379–403
- 963 Goldstein, A., Arzouan, Y., and Faust, M. (2012). Killing a novel metaphor and reviving a dead one: ERP
964 correlates of metaphor conventionalization. *Brain and Language* 123, 137–142
- 965 Gutiérrez, E. D., Shutova, E., Marghetis, T., and Bergen, B. K. (2016). Literal and metaphorical senses in
966 compositional distributional semantic models. In *Proceedings of the 54th Meeting of the Association*
967 *for Computational Linguistics* (Association for Computational Linguistics), 183–193. doi:10.18653/v1/
968 P16-1018
- 969 Guyon, I. and Elisseeff, A. (2003). An introduction to variable and feature selection. *J. Mach. Learn. Res.*
970 3, 1157–1182
- 971 Harris, Z. (1957). Co-occurrence and transformation in linguistic structure. *Language* 33, 283–340
- 972 Hill, F., Kiela, D., and Korhonen, A. (2013). Concreteness and corpora: A theoretical and practical analysis.
973 In *Proceedings of the Workshop on Cognitive Modeling and Computational Linguistics*. 75–83
- 974 Hill, F., Reichart, R., and Korhonen, A. (2015). Simlex-999: Evaluating semantic models with genuine
975 similarity estimation. *Computational Linguistics* 41, 665–695
- 976 Jacobs, A. M. and Kinder, A. (2017). “the brain is the prisoner of thought”: A machine-learning assisted
977 quantitative narrative analysis of literary metaphors for use in neurocognitive poetics. *Metaphor and*
978 *Symbol* 32, 139–160. doi:10.1080/10926488.2017.1338015
- 979 Jankowiak, K., Rataj, K., and Naskręcki, R. (2017). To electrify bilingualism: Electrophysiological insights
980 into bilingual metaphor comprehension. *PLoS ONE* 12, e0175578
- 981 Kartsaklis, D. and Sadzadeh, M. (2013). Prior disambiguation of word tensors for constructing sentence
982 vectors. In *Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing*.
983 1590–1601
- 984 Kiela, D. and Clark, S. (2014). A systematic study of semantic vector space model parameters. In
985 *Proceedings of the 2nd Workshop on Continuous Vector Space Models and their Compositionality*
986 *(CVSC) @ EACL 2014* (Gothenburg), 21–30
- 987 Kintsch, W. (2000). Metaphor comprehension: A computational theory. *Psychonomic Bulletin & Review* 7,
988 257–266. doi:10.3758/BF03212981
- 989 Kutas, M. and Federmeier, K. D. (2000). Electrophysiology reveals semantic memory use in language
990 comprehension. *Trends in cognitive sciences* 4, 463–470
- 991 Kutas, M. and Hillyard, S. A. (1984). Brain potentials during reading reflect word expectancy and semantic
992 association. *Nature* 307, 161

- 993 Lai, V. T. and Curran, T. (2013). ERP evidence for conceptual mappings and comparison processes during
994 the comprehension of conventional and novel metaphors. *Brain and Language* 127, 484–496
- 995 Lakoff, G. and Johnson, M. (1980). *Metaphors We Live By* (University of Chicago Press)
- 996 Levy, O. and Goldberg, Y. (2014a). Dependency-based word embeddings. In *Proceedings of the*
997 *52nd Annual Meeting of the Association for Computational Linguistics, ACL 2014, June 22-27, 2014,*
998 *Baltimore, MD, USA, Volume 2: Short Papers*. 302–308
- 999 Levy, O. and Goldberg, Y. (2014b). Neural word embedding as implicit matrix factorization. In *Proceedings*
1000 *of the 27th International Conference on Neural Information Processing Systems - Volume 2*. 2177–2185
- 1001 Martin, J. H. (1990). *A Computational Model of Metaphor Interpretation* (San Diego, CA, USA: Academic
1002 Press Professional, Inc.)
- 1003 McGregor, S., Agres, K., Purver, M., and Wiggins, G. (2015). From distributional semantics to conceptual
1004 spaces: A novel computational method for concept creation. *Journal of Artificial General Intelligence*
- 1005 McGregor, S., Jezek, E., Purver, M., and Wiggins, G. (2017). A geometric method for detecting semantic
1006 coercion. In *Proceedings of the 12th International Conference on Computational Semantics (IWCS)*
1007 (Montpellier: Association for Computational Linguistics)
- 1008 Mikolov, T., Chen, K., Corrado, G., and Dean, J. (2013). Efficient estimation of word representations in
1009 vector space. In *Proceedings of the International Conference on Learning Representations (ICLR)*
- 1010 Narayanan, S. (1999). Moving right along: A computational model of metaphoric reasoning about
1011 events. In *Proceedings of the Sixteenth National Conference on Artificial Intelligence and the Eleventh*
1012 *Innovative Applications of Artificial Intelligence Conference*. 121–127
- 1013 O'Brien, R. M. (2007). A caution regarding rules of thumb for variance inflation factors. *Quality &*
1014 *Quantity* 41, 673–690. doi:10.1007/s11135-006-9018-6
- 1015 Padó, S. and Lapata, M. (2007). Dependency-based construction of semantic space models. *Computational*
1016 *Linguistics* 33, 161–199
- 1017 Pennington, J., Socher, R., and Manning, C. D. (2014). GloVe: Global vectors for word representation. In
1018 *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*
1019 (Association for Computational Linguistics), 1532–1543. doi:10.3115/v1/D14-1162
- 1020 Rataj, K., Przekoracka-Krawczyk, A., and van der Lubbe, R. H. (2018). On understanding creative
1021 language: The late positive complex and novel metaphor comprehension. *Brain research* 1678, 231–244
- 1022 Reimer, M. (2001). Davidson on metaphor. *Midwest Studies in Philosophy* XXV, 142–155
- 1023 Rutter, B., Kröger, S., Hill, H., Windmann, S., Hermann, C., and Abraham, A. (2012). Can clouds dance?
1024 part 2: An erp investigation of passive conceptual expansion. *Brain and Cognition* 80, 301–310
- 1025 Sahlgren, M. (2008). The distributional hypothesis. *Italian Journal of Linguistics* 20, 33–53
- 1026 Schütze, H. (1992). Dimensions of meaning. In *Proceedings of the 1992 ACM/IEEE conference on*
1027 *Supercomputing*. 787–796
- 1028 Schütze, H. (1998). Automatic word sense discrimination. *Computational Linguistics* 24, 97–123
- 1029 Shutova, E. (2015). Design and evaluation of metaphor processing systems. *Computational Linguistics* 41
- 1030 Shutova, E., Kiela, D., and Maillard, J. (2016). Black holes and white rabbits: Metaphor identification with
1031 visual features. In *Proceedings of the 2016 Conference of the North American Chapter of the Association*
1032 *for Computational Linguistics: Human Language Technologies*. 160–170
- 1033 Shutova, E., Teufel, S., and Korhonen, A. (2012a). Statistical metaphor processing. *Computational*
1034 *Linguistics* 39, 301–353
- 1035 Shutova, E., van de Cruys, T., and Korhonen, A. (2012b). Unsupervised metaphor paraphrasing using a
1036 vector space model. In *Proceedings of COLING 2012: Posters*. 1121–1130

- 1037 Sperber, D. and Wilson, D. (2008). A deflationary account of metaphor. In *The Cambridge Handbook of*
1038 *Metaphor and Thought*, ed. R. Gibbs (Oxford University Press). 84–105
- 1039 Tsvetkov, Y., Boytsov, L., Gershman, A., Nyberg, E., and Dyer, C. (2014). Metaphor detection with cross-
1040 lingual model transfer. In *Proceedings of the 52nd Annual Meeting of the Association for Computational*
1041 *Linguistics (Volume 1: Long Papers)*. 248–258
- 1042 Utsumi, A. (2011). Computational exploration of metaphor comprehension processes using a semantic
1043 space model. *Cognitive Science* 35, 251–296. doi:10.1111/j.1551-6709.2010.01144.x
- 1044 Ward, T. B. (1994). Structured imagination: The role of category structure in exemplar generation.
1045 *Cognitive psychology* 27, 1–40
- 1046 Weiland, H., Bambini, V., and Schumacher, P. B. (2014). The role of literal meaning in figurative language
1047 comprehension: Evidence from masked priming erp. *Frontiers in Human Neuroscience* 8, 583
- 1048 Wolff, P. and Gentner, D. (2011). Structure-mapping in metaphor comprehension. *Cognitive Science* 35,
1049 1456–1488

In review

Figure 1.JPEG



In review



frontiers