Cognitive Systems Research 70 (2021) 101–108

Contents lists available at ScienceDirect

Cognitive Systems Research

journal homepage: www.elsevier.com/locate/cogsys

Reconstructing maps from text

Johnathan E. Avery *, Robert L. Goldstone, Michael N. Jones

Indiana University, Bloomington, United States

ARTICLE INFO

Keyword:
Semantic memory
Spatial cognition
Embodiment
Vector-space models

ABSTRACT

Previous research has demonstrated that Distributional Semantic Models (DSMs) are capable of reconstructing maps from news corpora (Louwerse & Zwaan, 2009) and novels (Louwerse & Benesh, 2012). The capacity for reproducing maps is surprising since DSMs notoriously lack perceptual grounding. In this paper we investigate the statistical sources required in language to infer maps, and the resulting constraints placed on mechanisms of semantic representation. Study 1 brings word co-occurrence under experimental control to demonstrate that standard DSMs cannot reproduce maps when word co-occurrence is uniform. Specifically, standard DSMs require that direct co-occurrences between city names in a corpus mirror the proximity between the city locations in the map in order to successfully reconstruct the spatial map. Study 2 presents an instance-based DSM that is capable of reconstructing maps independent of the frequency of co-occurrence of city names.

1. Introduction

Distributional Semantic Models (DSMs) posit cognitive mechanisms to explain how humans construct semantic representations for words from statistical regularities in natural language. Typically, these models represent words as points in a high-dimensional vector space, and similarity between words is measured as proximity in this semantic space. Semantic spaces are constructed by transforming natural language text corpora into vector spaces using a variety of mathematical techniques that span theoretically diverse learning and representation mechanisms (see Jones, Willits, & Dennis, 2015 for a review). In general, the relatively simple mechanisms used to construct DSMs have shown remarkable success at accounting for a broad range of semantic phenomena, such as human judgments of semantic similarity (Landauer & Dumais, 1997), semantic priming (Jones, Kintsch, & Mewhort, 2006), word association (Griffiths, Steyvers, & Tenenbaum, 2007), and analogies (Mikolov, Sutskever, Chen, Corrado, & Dean, 2013; Pennington, Socher, & Manning, 2014).

One major criticism of DSMs is that their representations are amodal and are not grounded in perception or action (Pecher & Zwaan, 2005; for a balanced discussion, see De Vega, Glenberg, & Graesser, 2012). Without perceptual grounding, DSMs may lack a necessary source of statistical information to fully represent semantic relationships between words. For instance, Glenberg and Robertson (2000) argue that without perceptual grounding, DSMs are incapable of predicting affordances of objects in novel situations, a task that humans can perform effortlessly.

Other researchers argue that a middle ground exists between amodal and fully grounded concept systems, based on redundancy and interdependence that exists between linguistic and perceptual information. Riordan and Jones (2011) demonstrated that many aspects of perceptual experiences are encoded in language by comparing similarity between spaces generated by DSMs and normed sets of features produced by human participants. Similarly, Roads and Love (2020) found significant overlap between lexically derived semantic spaces and the statistical distribution of objects in visual scenes. Transversely, Louwerse (2011) argues that conceptual relationships learned from language simultaneously inform and are shaped by the structure of embodied experience. Moreover, language has the capacity to encode embodied relations, so there may only be a need for symbols to be grounded intermittently (Louwerse, 2007, 2018).

Evidence demonstrating the capacities of DSMs to decode perceptual structures embedded in language comes from a body of work that reconstructs geographic relationships from text corpora. DSMs operate by embedding words in a high dimensional space by counting or predicting words in a given context (Baroni, Dinu, & Kruszewski, 2014). By placing words in a high dimensional space, similarity can be approximated as proximity. Louwerse and Zwaan (2009) used Latent Semantic Analysis (LSA; Landauer & Dumais, 1997) to create a semantic space of cities from large news corpora. By treating proximity in the high dimensional semantic space as a proxy for literal distance in physical space, Louwerse and Zwaan were then able to reproduce accurate maps of the USA. Similarly, Louwerse and Benesh (2012) demonstrated that even on the

* Corresponding author.
E-mail address: averjo@iu.edu (J.E. Avery).

https://doi.org/10.1016/j.cogsys.2021.07.007
Received 2 November 2020; Received in revised form 29 June 2021; Accepted 13 July 2021
Available online 8 August 2021

smaller corpus of The Lord of the Rings trilogy, LSA was able to closely reproduce a map of Middle Earth. Both instances demonstrated marked similarity to human-produced maps. Furthermore, this technique has been used to reproduce maps of various parts of the world using French (Louwerse, Cai, Hu, Ventura, & Jeuniaux, 2006), Chinese, and Arabic (Louwerse, Hutchinson, & Cai, 2012) corpora, as well as estimating the relative proximity between archeological sites using imprints on artifacts found in excavation (Recchia & Louwerse, 2016). This work demonstrates that the encoding of geographical location is not the result of a particular language, region, or historical period, but rather is endemic to how language is organized in order to encode features of perceptual experience (Louwerse, 2011).

The success at reproducing both real and fictional maps given corpora of text are fascinating demonstrations of the power of DSMs. However, the computational tools implemented in DSMs on natural language corpora obfuscate the statistical structures of language that lead to their success. On the one hand, using natural language corpora has the benefit of revealing what cultural knowledge has been encoded into language. On the other, there is an immense and diverse set of information encoded into language at small and large scales, and is not limited to geographic information (Semin, 2010).

In order to make any strong conclusions regarding the behavior and success of specific semantic models in reconstructing maps from text, the statistical properties of the training corpus must be taken under experimental control (Jones, Johns, & Recchia, 2012). It is possible to reproduce maps by counting first order (direct) co-occurrences between city names (Louwerse & Zwaan, 2009; Recchia & Louwerse, 2016), though the performance of this technique relies on using corpora that are many times the size required for LSA to be successful. Louwerse (2011) notes that, given a corpus of adequate size, first order relationships can perform analogously to LSA. It remains unclear whether the success demonstrated comes from the ability of LSA to use higher order co-occurrences (Landauer et al., 2007), or is directly related to the first order co-occurrences of city names in a text corpus.

We generate artificial corpora describing randomly generated maps to explore what statistical properties of a corpus enable DSMs to reproduce spatial distributions. When testing the capacities of DSMs to capture semantic information, researchers commonly use natural language corpora to generate high dimensional word representations (for example, Louwerse & Zwaan, 2009). While natural language corpora incorporate advantages with regard to external validity and the distributional statistics as they exist in common language usage, the sheer scale of natural language corpora do not necessarily clarify the mechanisms by which DSMs encode properties of language. For example, LSA is capable of modelling human data on a cued priming task when trained on a natural language corpus, but cannot replicate the task when the statistical attributes of the corpus are controlled (Jamieson et al., 2018, 2020). The ability of DSMs to produce satisfactory results at a given task stems not only from their computational techniques, but also from the statistical information that is embedded in the language itself (Johns, Jamieson, Crump, Jones, & Mewhort, 2020). It is not simply the modelling technique but also the corpus that is well suited to the task.

We bring the statistical properties of the corpus under experimental control by manipulating whether pairs of cities have a uniform or distance-based distribution of co-occurrences. In a distance-based corpus, a statement relating a pair of cities has a higher probability of occurring based on their relative distance, such that the probability of a statement relating two cities appearing in the corpus decreases as the distance between the two cities increases. In contrast, every statement has an equal probability of being sampled to produce a uniform-based corpus. We present two studies demonstrating the capacities of various DSMs at reproducing maps. Study 1 implements a set of standard DSMs, while Study 2 implements a model that theoretically diverges from standard DSMs, demonstrating qualitative differences between the DSMs evaluated.

2. Methods

Our goal is to determine whether a DSM is able to reproduce a map from systematically constructed linguistic descriptions of the relationships between cities. Here, we briefly describe our steps for evaluating DSM performance, with a more in-depth description to follow. We start by generating a corpus of descriptions of a map. First, we randomly generate a set of maps of varying distributions of artificial "cities." Next, for every pair of cities, we generate statements that describe the relationships between the two cities. In particular, we use two sets of relationships: North, South, East, and West; and near and far. We sample the sets of descriptions either uniformly or based on distance in order to yield multiple corpora. Once the corpora have been generated, we test a set of DSMs for their ability to reproduce the original maps based on the linguistic descriptions of the relationships between cities. We follow an outline of the steps used in Louwerse and Zwaan (2009) and Louwerse and Benesh (2012) in order to move from a corpus of text to a two-dimensional map, with minor modifications to their process. In particular, we start by training a model on a given corpus. We then find the cosine similarity between the vectors representing each pair of cities, and convert the similarity into distance using Shepard (1987) exponential law of generalization. We convert the distance matrix into a two-dimensional plot using multidimensional scaling. Finally, we compare the two-dimensional plot yielded by the DSMs to the original map using bidimensional regression (Friedman & Kohler, 2003). Bidimensional regression is a special measure designed to compare two-dimensional plots that yields a measure of goodness of fit. Study 1 evaluates the performance of standard DSMs at reproducing maps given corpora where statements relating cities are sampled either uniformly or based on distance. Study 2 implements ITS coupled with a context retrieval mechanism, and evaluates this model's performance at reproducing maps as a function of sampling.

2.1. Generate maps

We generated three maps with different distributions of 'cities': random, clustered, and circular. The maps are displayed in the first column of Fig. 1. There are 20 cities on each map, labelled with the letters 'A' through 'T'. We chose these three distributions of cities to vary the amount of external distribution information available to disambiguate the cities, and cover a range of challenges.

2.2. Generate corpora

We generated multiple sets of corpora for each combination of conditions. The conditions include the distribution of cities (random, clustered, shape), the description set (either North/South/East/West, or near/far), and the type of sampling (uniform or distance-based). In total, there were 12 corpora.

In order to generate the corpora, we created statements about each pair of cities on a given map. The statements took the form: [city] [relationship to] [city]. We used two sets of relationships: near/far, and North/South/East/West. Sections 2.2.1 and 2.2.2 explain how these relationships are operationalized; see Fig. 2 for additional visualization of the process.

2.2.1. Near/Far

For each map, we found the average distance between cities. Subsequently, when generating a statement relating two cities (e.g. cities 'A' and 'B'), if the distance between the two cities was above the average distance we would label the relationship between the two cities as 'far' (e.g. 'A is far from B'). Conversely, if the distance between the two cities was lower than the average distance between all cities, we would label the relationship between the two cities as 'near' (e.g. 'A is near to B').

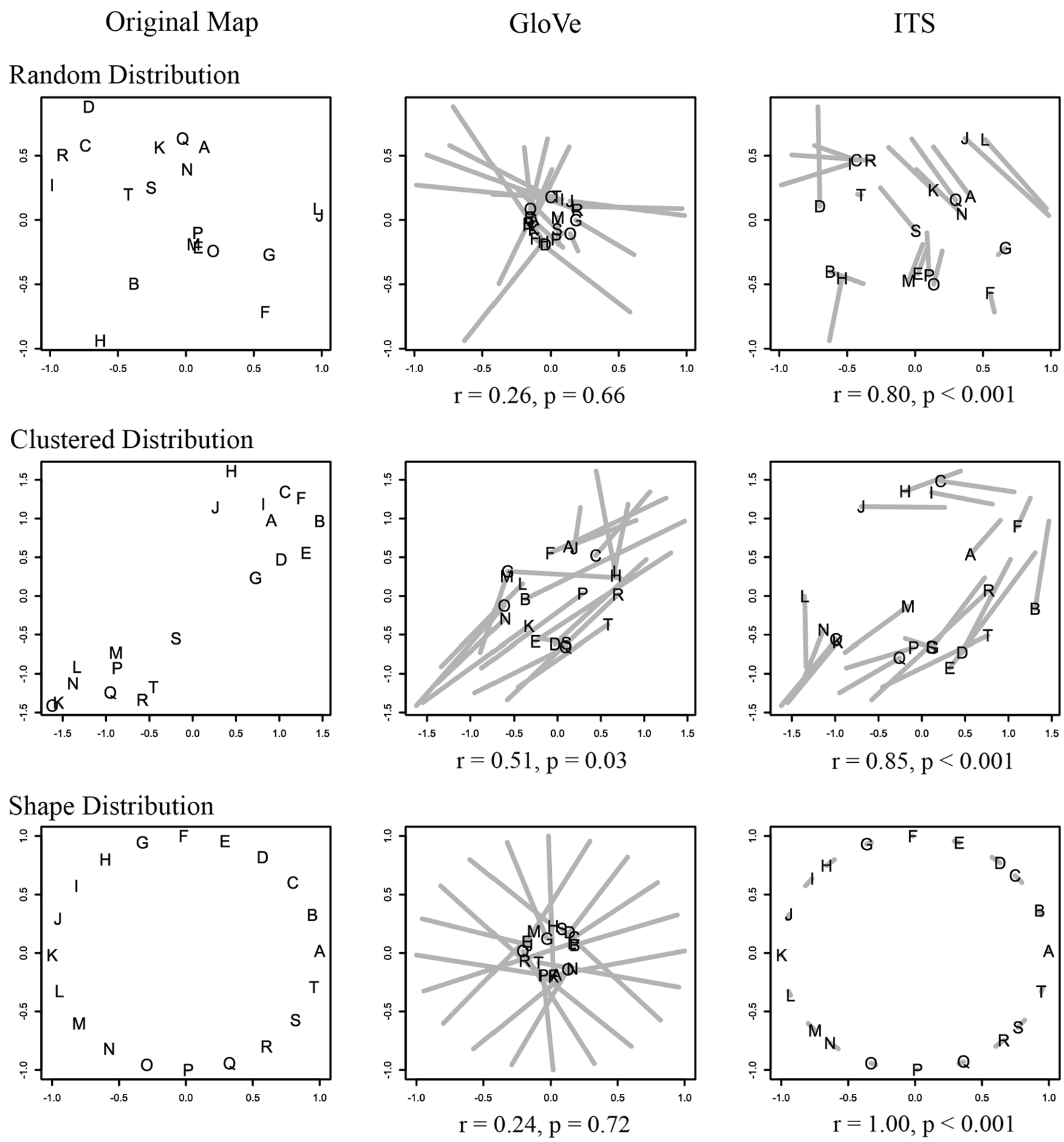

Fig. 1. A representative sample of the performance of models at reconstructing the original map given the North/South/East/West relationship set with uniform sampling. From left to right, the columns present 1) the original maps, 2) the maps produced by GloVe (Study 1), and 3) the maps produced by ITS (Study 2). Each row shows one of the three distributions tested. The r coefficient and significance is noted below each map. The predicted distributions were rotated and scaled to minimize their difference from the original map using the Procrustes algorithm in R (Oksanen et al., 2020). The grey lines indicate the offset from a predicted location and the location in the original distribution.

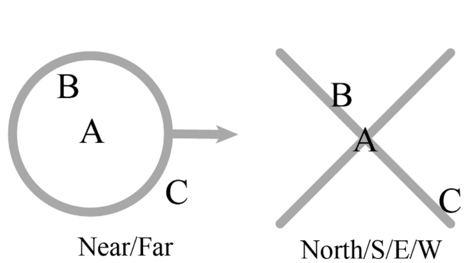

Fig. 2. Visualization of how relationships between cities were derived. On the left, the grey circle denotes the Near/Far boundary, and yields the statements, “B is Near A”, and “C is far from A”. On the right, the grey lines denote the North, South, East, and West quadrants, yielding the statements, “B is North of A”, and “C is East of A”.

2.2.2. North/South/East/West

In order to formalize the directional relationship between cities, we divided the area surrounding a city into four quadrants, rotated by 45 degrees. For example, if some city ‘B’ appeared in the quadrant above a city ‘A’, it would be labeled as ‘north of’ city ‘A’, thereby generating the statement ‘B is North of A’.

2.2.3. Generating the statements

We compared each city to every other city in order to generate a sentence regarding the two cities’ relationships. Therefore, there are two sentences relating any two cities – with the relationship reversed. If the relationship between cities A and B is that ‘A is North of B’, then there will necessarily be a converse statement in the corpus that ‘B is South of A’. There are no statements relating a city to itself (e.g. ‘A is South of A’). In the statements, the phrases were condensed such that the relationship was just one word (e.g. ‘is North of’ was reduced to ‘north_of’).

For each set of relationships, there are 380 possible unique statements relating all of the cities to one another. We created a statement set for both of the relationship sets; additionally we created a third statement set by combining the other two statement sets.

2.2.4. Sampling the statements

A corpus was generated by sampling a statement set either uniformly or based on distance. Each corpus consisted of 10,000 statements. If the set of statements was sampled uniformly, then each statement occurred an equal number of times in the final corpus. If the set of statements was sampled based on distance, statements relating cities that were closer together were sampled more frequently than statements relating cities that were further away. The probability of sampling a statement t_{ij} that relates cities i and j is based on the distance between the two cities and is given by

$$P(t_{ij}) = \frac{\max(d) + 1 - d_{ij}}{\sum_i^I \sum_j^J \max(d) + 1 - d_{ij}} \tag{1}$$

where d_{ij} is the distance between cities i and j, and max(d) is the maximum distance between two cities.

2.3. Vector spaces produced by DSMs

We trained a set of models on each corpus. The set of models includes LSA (Landauer, Foltz, & Laham, 1998), Continuous Bag Of Words (CBOW; Mikolov et al., 2013), and Global Vectors for word representation (GloVe; Pennington et al., 2014), Positive Pointwise Mutual Information (PPMI; Bullinaria & Levy, 2007), – to be described in further detail in Study 1 – and the exemplar model, ITS (Jamieson, Avery, Johns, & Jones, 2018) – to be described in further detail in Study 2.

Each of these models – with the exception of PPMI – operates by building a vector space, where words are placed in an arbitrary high-dimensional space. Their location in that space determines their similarity to the other words in that space, such that words that are near to each other are similar, while distant words are dissimilar. The cosine between two vectors is a commonly used metric to compute word similarities (Bullinaria & Levy, 2007).

Rather than working with similarities, the steps that follow require distances. We transform the cosine similarity matrix into a distance matrix using Shepard (1987) universal law of generalization, such that the distance d between two word vectors i and j is given by

$$d_{ij} = \exp(-\gamma sim(i,j)) \tag{2}$$

where $sim(i, j)$ is the similarity between two word vectors i and j, and γ is a monotonic scaling factor.

The cosine similarities were computed between each city, and the similarity matrix was transformed into a distance matrix. Our processes yielded a distance matrix for each model, for each sampling method, for each description set, for each map, for a total of 90 distances matrices for evaluation.

2.4. Generating and evaluating the map

We used multidimensional scaling (MDS) to generate a 2-dimensional map from the distance matrix. We then used bidimensional regression to evaluate the map produced by the model against the original map.

2.4.1. MDS

MDS is a well-established technique to transform a distance matrix into a plot of some arbitrary dimensionality (Kruskal & Wish, 1978). We used Kruskal's non-metric MDS as implemented in R (MASS; Venables & Ripley, 2002). Non-metric MDS produces a solution that matches the ordinal ranking of the distances provided in the distance matrix. The solutions produced by non-metric MDS are subject to rotations, shifts, scaling, and flips.

2.4.2. Bidimensional regression

We used bidimensional regression to evaluate how well each MDS solution recreated the map from which it was derived (Friedman & Kohler, 2003; Louwerse & Zwaan, 2009). Bidimensional Regression is a measure of how well two maps align. The measure of interest for this study is the r coefficient. The r coefficient indicates how well two spatial distributions of points match each other. We generated our r coefficient using the affine bidimensional regression in R (BiDimRegression; Carbon, 2013).

3. Study 1 – Standard DSMs

The purpose of Study 1 is to explore under what conditions standard DSMs can and cannot produce maps as demonstrated by Louwerse and Zwaan (2009) and Louwerse and Benesh (2012). We aimed to test the hypothesis that the first order co-occurrence of words drives the performance of standard DSMs. Accordingly, we manipulated the frequency by which statements are sampled.

3.1. DSMs for evaluation

We first evaluate the ability of four state-of-the-art DSMs to produce spatial distributions from linguistic descriptions of maps[1]. While the set of models used here is not exhaustive, it is representative of the range of techniques of the best-performing models used in the field of semantic memory.

3.1.1. LSA

LSA is a well-established method of creating a semantic space by applying a combination of term frequency weightings and singular value decomposition (SVD) to a word-by-document frequency count of words in a text corpus (Landauer et al., 1998). SVD is a linear algebra technique that smooths a given matrix on the basis of its eigenvalues, or principle sources of variance. We used the python package, gensim (Řehůřek & Sojka, 2010), to implement LSA.

3.1.2. CBOW

CBOW is a neural network implementation of a DSM within the word2vec architecture (Mikolov et al., 2013), where the localist input layer is a given word, and the output layer is the set of words in whose context the input word is found, with a single hidden layer. The algorithm finds the appropriate weights between nodes by minimizing the error between the context that the model predicts and the actual context in which a word occurs. The weights between the input and hidden layers are treated as word embeddings in a high dimensional vector space. We used the CBOW implementation provided by the Python gensim package, with 50 training iterations, and the standard 300 nodes in the hidden layer.

3.1.3. GloVe

GloVe generates word vector representations using regression to fit an arbitrarily high dimensional vector space to a word co-occurrence measure (Pennington et al., 2014). The technique minimizes difference between distributions between the co-occurrence matrix and the vector space of arbitrarily high dimensionality. We used the GloVe implementation provided by Pennington et al., with 50 dimensions and a maximum 15 iterations to convergence.

3.1.4. PPMI

PMI is a log transform of the conditional probabilities of the co-occurrence of words (Bullinaria & Levy, 2007). Positive PMI (PPMI) sets negative PMI values to zero. PPMI is distinct from the other DSMs, as it produces a value relating two words based on their conditional probability, skipping the step in vector space models in which a word is

[1] Code for the analyses presented here available at: https://github.com/mastercciolі/reconstructing-space-from-text

placed in a high dimensional space and the relationship between words is derived from their cosine similarity. We used the PPMI implementation provided in the Python gensim package.

3.2. Results

Table 1 presents the performance of each model at reproducing the original map given the different corpus sampling conditions for each relationship set. The significance of how well the map is reconstructed is indicated with asterisks. When the corpus is generated using uniform sampling, only a few models produced maps that significantly reproduced the original map. In contrast, when a corpus is generated with distance-based sampling, all models produce significant reproductions of the maps. The combined corpus yielded no qualitative difference in the models' performance.

The center column in Fig. 1 provides a visual demonstration of the performance of GloVe at reproducing the maps given the North/South/East/West relationship set with uniform sampling. Notably, the reproduction of the two clusters by GloVe demonstrates that, although the reproduction is significant, a value of $r = 0.77$ corresponds to only a rough approximation of the original space. Note how in this example, the two clusters are not well separated. Overall, models tested on corpora generated by uniform sampling are mediocre at best at reconstructing the maps.

3.3. Discussion

The manipulation of the frequency by which statements are sampled illuminates the strength of standard DSMs at exploiting first order co-occurrences to reproduce maps from text. Generally, objects that are related to one another are more likely to be discussed within the same context. This principle extends to the spatial locations of cities, such that cities that are close to one another are more likely to be discussed in the same context than cities that are far apart. Louwerse (2011) comments that first order co-occurrence statistics would likely obviate the need for LSA given text corpora of sufficiently large size, however that the strength of LSA emerges from its ability to extrapolate relationships in corpora that are orders of magnitude smaller than would be required for first order statistics. In this study, there are a small number of statements in comparison to the size of the corpus, which likely leads to the observed success of PPMI in our tests. However, when we control for the frequency of co-occurrence, standard DSMs are not able to accurately co-locate cities in semantic space, indicating that some measure of first order co-occurrence must be available in order for standard DSMs to perform the map reconstruction task.

4. Study 2 – A retrieval-based DSM

Study 1 demonstrates that standard DSMs cannot reconstruct accurate spatial relationships when first order co-occurrences are uniform.. Specifically, standard DSMs treat frequently co-occurring words as more similar to one another. However, the structure of the corpora we generated seemed likely to contain sufficient information to reproduce the distributions regardless of sampling condition.

A drawback of standard DSMs is that the representation of a given word is reduced to a single point in semantic space. We hypothesized that in order for a model to reproduce the maps in the uniform sampling condition, the model would need the ability to flexibly adapt relationships between words. We use the Instance Theory of Semantics (ITS; Jamieson et al., 2018) because it flexibly modifies word representations during the retrieval process, rather than during the learning process (Jones, 2019). Study 2 demonstrates that a model with a sufficiently flexible learning, storage, and retrieval process is capable of learning the spatial distributions of the three maps from linguistic descriptions independent of the two sampling procedures used to generate the corpora.

4.1. Model

ITS uses a multiple-trace episodic memory store. Each instance is a set of words that co-occur in the same context, and is uniquely stored as a trace in memory. Memory may then be probed in order to retrieve an abstracted semantic vector representation of the word. When memory is probed, traces that contain the probe are recalled and combined into an echo. The echo is the normalized sum of all the contexts in which the probe word occurred (cf. Hintzman, 1986). When the probe is composed of multiple words, the echo is composed only of contexts where both words occur.

An advantage of using an abstraction-at-retrieval model lies in the flexibility of the retrieval process (Crump, Jamieson, Johns, & Jones, 2020). Specifically, ITS can take context of retrieval into account, a key drawback of standard DSMs noted by numerous researchers (Jamieson et al., 2018; Jones, 2019; Kintsch, 2000). We modify the retrieval process in two ways in order to maximize ITS ability to vary word meanings based on context. First, we modify the echo to simply yield the context of the probe without the probe itself. Second, we modify the process by which similarity is evaluated.

4.1.1. Context of probe

In ITS, the retrieved echo is composed of the vectors representing the set of words that co-occur with the probe, as well as the probe itself. Here, we are interested in comparing the contexts in which words occur. Given that the echo contains both the probe and the context of the probe, we must separate the context from the probe. Specifically, we define the

Table 1
Presents the r coefficients for each model at reproducing a map distribution given a particular sampling procedure and relationship set. The Near/Far relationship set is denoted 'N/F', while the North/South/East/West relationship set is denoted 'N/S/E/W'.

		Distance			Uniform		
Model	**Map Distribution**	**N/F**	**N/S/E/W**	**Combined**	**N/F**	**N/S/E/W**	**Combined**
LSA	Random	0.79***	0.87***	0.89***	0.21	0.27	0.30
	Clustered	0.75***	0.75***	0.90***	0.21	0.45	0.18
	Circular	0.31	0.74***	0.99***	0.14	0.28	0.21
CBOW	Random	0.55*	0.60**	0.45	0.37	0.37	0.36
	Clustered	0.90***	0.86***	0.88***	0.55*	0.46	0.48
	Circular	0.98***	0.98***	0.98***	0.29	0.12	0.48
GloVe	Random	0.85***	0.90***	0.94***	0.17	0.26	0.13
	Clustered	0.97***	0.96***	0.96***	0.27	0.51*	0.31
	Circular	1.00***	0.99***	0.99***	0.16	0.24	0.39
PPMI	Random	0.94***	0.93***	0.91***	0.16	0.16	0.24
	Clustered	0.97***	0.97***	0.97***	0.30	0.45	0.19
	Circular	1.00***	1.00***	1.00***	0.13	0.29	0.39

$p < 0.05$ - *; $p < 0.01$ - **; $p < 0.001$ - ***.

context of a probe as the echo of a probe without the probe itself. Formally,

$$cont(probe) = echo(probe) - probe \quad (3)$$

where *cont* denotes the context, and *echo(probe)* yields the echo given a probe as defined by ITS.

4.1.2. Retrieval process

In general, the similarity between words is treated as a direct cosine comparison between the vectors representing two words. Here, we deviate from convention by treating similarity as the comparison between the two words and some set of tertiary words. That is, the similarity between two words A and B can be approximated as what is shared between the context of word A and some other word C and the context of word B and that same word C. Words A and B are similar to the extent that the features that constitute the context of words A and C are shared with the features that constitute the context of words B and C. This definition of similarity does not deviate from the distributional hypothesis, but rather serves as an alternate formalization.

We define the similarity *sim* between two words *a* and *b* as

$$sim(a,b) = \sum_{i}^{I} cosine(cont(a,i), cont(b,i)) \quad (4)$$

where *I* is the set of all unique words used in the corpus, *cosine* is the cosine similarity.

Here we present a concrete example of how such a retrieval process in ITS might elicit a spatial distribution of cities from a linguistic description of their relationships. Consider three cities A, B, and C, such that A and B are near each other and are both far from C. When comparing A and B, we use all the words in the corpus except A and B. For instance, we want to compare the *cont(A, C)* with *cont(B, C)*. The *cont(A, C)* yields 'far_from', likewise the *cont(B, C)* also yields 'far_from'. Therefore, when evaluating *cosine(cont(A, C), cont(B, C))*, the yielded value is high because the context for both pairs is identical.

In contrast, if we want to compare cities A and C, we would use B as a tertiary word of comparison. In this case, *cosine(cont(A, B), cont(C, B))* would yield a low similarity because the *cont(A, B)* yields 'near_to' while the *cont(C, B)* yields 'far_from'. Thusly, the process of eliciting the similarity between two words via the shared contexts with tertiary words can accurately elicit spatial locations from linguistic descriptions.

4.1.3. Comparing retrieval processes between DSMs

The retrieval process presented in the previous section is an expansion of that used in Jamieson et al. (2018), and is different than what is typically applied to DSMs. While it is possible to use the standard DSM similarity process with this model, the process would yield the same results as standard DSMs. Specifically, the standard similarity process would yield an overt sensitivity to first order co-occurrence statistics, and no ability to respond to higher order statistics in the absence of first order statistics. Such a failing reveals the necessity of a modified retrieval mechanism.

Various techniques have been applied to standard DSM vector representations to accommodate a variety of tasks, such as metaphor comprehension (Kintsch, 2000), hypernym relationships (Lenci & Benotto, 2012), analogy interpretation (Mikolov et al., 2013), and asymmetric relationships (Kintsch, 2014). However, applying a modified similarity process to the vector representations produced by standard DSMs in the uniform sampling condition would not yield meaningful results, given that the cosine similarities between cities are randomly distal from one another. Study 2 is meant to demonstrate the need for a fundamental shift to access higher order relationships between words when first order co-occurrence statistics do not provide useful information for relating cities.

4.2. Results

Table 2 presents the performance of ITS at reproducing the original map given the different corpus sampling conditions for each relationship set. The significance of how well the map is reconstructed is indicated with asterisks. Independent of sampling condition, ITS is able to produce significant reconstruction of all original maps. Notably, the lowest performing reconstruction has an $r = 0.80$, well above the highest performing model in Study 1 in the uniform sampling condition.

The right column in Fig. 1 provides a visual demonstration of the performance of ITS at reproducing the maps given the North/South/East/West relationship set with uniform sampling. The ITS reconstruction of the shape provides compelling visual evidence that DSMs are capable of reproducing maps if the DSM represents each word exhaustively in terms of the lexical contexts in which it occurs.

5. General discussion

Demonstrations by Louwerse and Zwaan (2009) and Louwerse and Benesh (2012) show that spatial distributions can be elicited from text. Given that amodal DSMs are not grounded by any perceptual input, it is surprising that DSMs can reproduce spatial distributions whatsoever. Here we bring sampling frequency under experimental control to demonstrate that frequency of co-occurrence provides sufficient information to enable standard DSMs to reproduce spatial distributions.

Study 1 demonstrates that standard "abstraction-at-learning" DSMs are only able to successfully reconstruct spatial distributions in the condition when linguistic descriptions are sampled based on proximity. When cities that are near each other are discussed more frequently than cities that are far apart, standard DSMs are able to reproduce their spatial distributions. When cities are discussed with uniform frequency, standard DSMs are not able to reproduce spatial distributions. The inability of standard DSMs to reproduce maps in the uniform sampling condition may indicate that this class of models uses first order word associations in order to bootstrap to second order, semantic word relationships (Kelly, Ghafurian, West, & Reitter, 2020).

Study 2 explores a cognitively inspired 'abstraction-at-retrieval' DSM, where the semantic relationship between two words is dependent on the context in which the words co-occur, unlike 'abstraction-at-learning' DSMs. In an instance-based DSM, there is no stored semantic memory, only episodic memory. Semantic representations are constructed on-the-fly as a product of the episodic retrieval mechanism in response to an environmental probe. The model presented in Study 2 demonstrates that an instance-based DSM is capable of reproducing spatial distributions given uniformly sampled descriptions of cities. The instance-based DSM does not require first order word co-occurrences in order to bootstrap second order, semantic relationships between words.

Both Studies 1 and 2 replicate and expand upon previous demonstrations that spatial information can be successfully reconstructed by DSMs even when they are not biased with information about inter-object distances (Louwerse & Benesh, 2012; Louwerse & Zwaan, 2009; Recchia

Table 2
Presents the r coefficients for ITS at reproducing a map distribution given a particular sampling procedure and relationship set.

Corpus	Distribution	Distance	Uniform
N/F	Random	0.79***	0.79***
	Clustered	0.96***	0.96***
	Circular	1.00***	1.00***
N/S/E/W	Random	0.63**	0.80***
	Clustered	0.87***	0.85***
	Circular	1.00***	1.00***
Combined	Random	0.81***	0.81***
	Clustered	0.96***	0.96***
	Circular	1.00***	1.00***

$p < 0.05$ - *; $p < 0.01$ - **; $p < 0.001$ - ***.

& Louwerse, 2014). Standard DSMs are capable of producing spatial distributions with distance-based sampling, which may exist in natural language corpora. However, requiring distance-based information threatens the generalizability of standard DSMs in representing semantic information. If alternatives like ITS are available that do not require biased samples, then they may offer more robust solutions to the problem of extracting structural patterns from text. Study 2 extends the representation of space toward a comprehension process, whereby spatial distributions can be elicited by DSMs independent of the first order co-occurrence of cities.

We have presented an instance-based model that is simple in its instantiation and applied to a constrained problem. Further explorations with this model could address the computational complexity of the model, as well as the applicability of the model to more naturalistic problems.

There is a tradeoff between the flexibility of the instance-based DSM and computational cost. In standard DSMs, the initial training is computationally expensive, and the subsequent comparison of word meanings is computationally cheap. Our instance-based DSM makes the opposite tradeoff, and future research into the implementation of the algorithm should aim to reduce the computational complexity at the point of retrieval.

We implemented the instance-based model in a highly artificial setting. While the limitations of standard DSMs are demonstrated in Study 1, standard DSMs have been useful in a variety of domains related to natural language phenomena. Given our demonstrations here, it is unclear whether our instance-based model will provide an advantage given more naturalistic language that is not controlled for uniformity of co-occurrence. Future research would usefully investigate whether any such advantage exists for our model.

The instance-based model in Study 2 provides a clear advantage over traditional DSMs. Standard DSMs do not use higher order information in the uniformly sampled corpus (as evidenced by Study 1) even though that information suffices to allow an instance-based model to recreate the maps used to generate text inputs (as evidenced by Study 2). Standard DSMs capitalize on the co-occurrence between words in language. Instance-based models take advantage of the conditional relationships between words. The future of semantic modelling requires a need to find a balance between the computational efficiency of standard DSMs and the flexibility of instance-based models.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

Baroni, M., Dinu, G., & Kruszewski, G. (2014). Don't count, predict! A systematic comparison of context-counting vs. Context-predicting semantic vectors. In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)* (pp. 238–247).

Bullinaria, J. A., & Levy, J. P. (2007). Extracting semantic representations from word co-occurrence statistics: A computational study. *Behavior Research Methods, 39*(3), 510–526. https://doi.org/10.3758/BF03193020.

Carbon, C.-C. (2013). BiDimRegression: Bidimensional regression modeling using R. *Journal of Statistical Software, 52*(Code Snippet 1). https://doi.org/10.18637/jss.v052.c01.

Crump, M. J. C., Jamieson, R. K., Johns, B. T., & Jones, M. N. (2020). Controlling the retrieval of general vs specific semantic knowledge in the instance theory of semantic memory. *Proceedings of the 42nd Annual Meeting of the Cognitive Science Society*.

De Vega, M., Glenberg, A., & Graesser, A. (2012). *Symbols and embodiment: Debates on meaning and cognition*. Oxford University Press.

Friedman, A., & Kohler, B. (2003). Bidimensional regression: Assessing the configural similarity and accuracy of cognitive maps and other two-dimensional data sets. *Psychological Methods, 8*(4), 468–491. https://doi.org/10.1037/1082-989X.8.4.468.

Glenberg, A. M., & Robertson, D. A. (2000). Symbol grounding and meaning: A comparison of high-dimensional and embodied theories of meaning. *Journal of Memory and Language, 43*(3), 379–401.

Griffiths, T. L., Steyvers, M., & Tenenbaum, J. B. (2007). Topics in semantic representation. *Psychological Review, 114*(2), 211–244. https://doi.org/10.1037/0033-295X.114.2.211.

Hintzman, D. L. (1986). "Schema abstraction" in a multiple-trace memory model. *Psychological Review, 93*(4), 411–428. https://doi.org/10.1037/0033-295X.93.4.411.

Jamieson, R. K., Avery, J. E., Johns, B. T., & Jones, M. N. (2018). An instance theory of semantic memory. *Computational Brain & Behavior, 1*(2), 119–136. https://doi.org/10.1007/s42113-018-0008-2.

Jamieson, R. K., Avery, J. E., Johns, B. T., & Jones, M. N. (2020). Correction to: An instance theory of semantic memory. *Computational Brain & Behavior, 3*(1), 126–127. https://doi.org/10.1007/s42113-019-00046-x.

Johns, B. T., Jamieson, R. K., Crump, M. J. C., Jones, M. N., & Mewhort, D. J. K. (2020). Production without rules: Using an instance memory model to exploit structure in natural language. *Journal of Memory and Language, 115*, 104165. https://doi.org/10.1016/j.jml.2020.104165.

Jones, M. N. (2019). When does abstraction occur in semantic memory: Insights from distributional models. *Language, Cognition and Neuroscience, 34*(10), 1338–1346. https://doi.org/10.1080/23273798.2018.1431679.

Jones, M. N., Johns, B. T., & Recchia, G. (2012). The role of semantic diversity in lexical organization. *Canadian Journal of Experimental Psychology/Revue Canadienne de Psychologie Expérimentale, 66*(2), 115–124. https://doi.org/10.1037/a0026727.

Jones, M. N., Kintsch, W., & Mewhort, D. J. K. (2006). High-dimensional semantic space accounts of priming. *Journal of Memory and Language, 55*(4), 534–552. https://doi.org/10.1016/j.jml.2006.07.003.

Jones, M. N., Willits, J., & Dennis, S. (2015). Models of semantic memory. *Oxford Handbook of Mathematical and Computational Psychology*, 232–254.

Kelly, M. A., Ghafurian, M., West, R. L., & Reitter, D. (2020). Indirect associations in learning semantic and syntactic lexical relationships. *Journal of Memory and Language, 115*, 104153. https://doi.org/10.1016/j.jml.2020.104153.

Kintsch, W. (2000). Metaphor comprehension: A computational theory. *Psychonomic Bulletin & Review, 7*(2), 257–266. https://doi.org/10.3758/BF03212981.

Kintsch, W. (2014). Similarity as a function of semantic distance and amount of knowledge. *Psychological Review, 121*(3), 559–561. https://doi.org/10.1037/a0037017.

Kruskal, J. B., & Wish, M. (1978). *Multidimensional scaling*. Sage.

Landauer, T. K., & Dumais, S. T. (1997). A solution to Plato's problem: The latent semantic analysis theory of acquisition, induction, and representation of knowledge. *Psychological Review, 104*(2), 211–240. https://doi.org/10.1037/0033-295X.104.2.211.

Landauer, T. K., Foltz, P. W., & Laham, D. (1998). An introduction to latent semantic analysis. *Discourse Processes, 25*(2–3), 259–284. https://doi.org/10.1080/01638539809545028.

Landauer, T. K., McNamara, D. S., Dennis, S., & Kintsch, W. (Eds.). (2007). *Handbook of latent semantic analysis*. Erlbaum.

Lenci, A., & Benotto, G. (2012). Identifying hypernyms in distributional semantic spaces. In SEM 2012: The First Joint Conference on Lexical and Computational Semantics – Volume 1: Proceedings of the Main Conference and the Shared Task, and Volume 2: Proceedings of the Sixth International Workshop on Semantic Evaluation (SemEval 2012) (pp. 75–79). https://www.aclweb.org/anthology/S12-1012.

Louwerse, M. (2007). Symbolic or embodied representations: A case for symbol interdependency. *Handbook of Latent Semantic Analysis, 107*, 120.

Louwerse, M. (2011). Symbol interdependency in symbolic and embodied cognition. *Topics in Cognitive Science, 3*(2), 273–302. https://doi.org/10.1111/j.1756-8765.2010.01106.x.

Louwerse, M. (2018). Knowing the meaning of a word by the linguistic and perceptual company it keeps. *Topics in Cognitive Science, 10*(3), 573–589. https://doi.org/10.1111/tops.12349.

Louwerse, M., & Benesh, N. (2012). Representing spatial structure through maps and language: Lord of the rings encodes the spatial structure of middle earth. *Cognitive Science, 36*(8), 1556–1569. https://doi.org/10.1111/cogs.12000.

Louwerse, M., Cai, Z., Hu, X., Ventura, M., & Jeuniaux, P. (2006). Cognitively inspired NLP-based knowledge representations: Further explorations of Latent Semantic Analysis. *International Journal on Artificial Intelligence Tools, 15*(06), 1021–1039.

Louwerse, M., Hutchinson, S., & Cai, Z. (2012). The Chinese route argument: Predicting the longitude and latitude of cities in China and the Middle East using statistical linguistic frequencies. *Proceedings of the Annual Meeting of the Cognitive Science Society*, 34(34).

Louwerse, M., & Zwaan, R. A. (2009). Language encodes geographical information. *Cognitive Science, 33*(1), 51–73. https://doi.org/10.1111/j.1551-6709.2008.01003.x.

Mikolov, T., Sutskever, I., Chen, K., Corrado, G. S., & Dean, J. (2013). Distributed representations of words and phrases and their compositionality. *Advances in Neural Information Processing Systems*, 3111–3119.

Oksanen, J., Blanchet, F.G., Kindt, R., Legendre, P., Minchin, P.R., O'Hara, R.B., Simpson, G.L., Solymos, P., Stevens, M.H.H., Wagner, H. (2020). vegan: Community Ecology Package. R package version 2.5-7, URL https://CRAN.R-project.org/package=vegan.

Pecher, D., & Zwaan, R. A. (Eds.). (2005). *Grounding cognition: The role of perception and action in memory, language, and thinking*. Cambridge University Press.

Pennington, J., Socher, R., & Manning, C. D. (2014). Glove: Global vectors for word representation. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)* (pp. 1532–1543).

Recchia, G., & Louwerse, M. (2014). Grounding the ungrounded: Estimating locations of unknown place names from linguistic associations and grounded representations. *Proceedings of the Annual Meeting of the Cognitive Science Society*, 36(36).

Recchia, G. L., & Louwerse, M. M. (2016). Archaeology through computational linguistics: Inscription statistics predict excavation sites of indus valley artifacts. *Cognitive Science, 40*(8), 2065–2080. https://doi.org/10.1111/cogs.2016.40.issue-810.1111/cogs.12311.

Řehůřek, R., & Sojka, P. (2010). Software framework for topic modelling with large corpora. In *Proceedings of the LREC 2010 Workshop on New Challenges for NLP Frameworks* (pp. 45–50).

Riordan, B., & Jones, M. N. (2011). Redundancy in perceptual and linguistic experience: Comparing feature-based and distributional models of semantic representation. *Topics in Cognitive Science, 3*(2), 303–345. https://doi.org/10.1111/j.1756-8765.2010.01111.x.

Roads, B. D., & Love, B. C. (2020). Learning as the unsupervised alignment of conceptual systems. *Nature Machine Intelligence, 2*(1), 76–82. https://doi.org/10.1038/s42256-019-0132-2.

Semin, G. R. (2010). Culturally situated linguistic ecologies and language use: Cultural tools at the service of representing and shaping situated realities. In M. J. Gelfand, C. Chiu, & Y. Hong (Eds.), *Advances in Culture and Psychology*. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780195380392.001.0001.

Shepard, R. (1987). Toward a universal law of generalization for psychological science. *Science, 237*(4820), 1317–1323. https://doi.org/10.1126/science.3629243.

Venables, W. N., & Ripley, B. D. (2002). *Modern Applied Statistics with S (Fourth)*. Springer. http://www.stats.ox.ac.uk/pub/MASS4.