Title: Topology-preserving topographic representation for implementation of linguistic intentionality

Author: Moisl, H. (2025)Topology-Preserving Topographic Representation for Implementation of Linguistic Intentionality, Journal of Quantitative Linguistics
32, https://doi.org/10.1080/09296174.2025.2583941.

Keywords: Linguistic meaning, intentionality, topology preservation, topographic representation, artificial neural network, self-organizing map, topology representing network

Abstract

A tradition in western thought ranging from Aristotle to theories of mental content in present-day linguistics and cognitive science says that the meaning of a word is its signification of a mental concept, and a mental concept is a representation of the mind-external environment causally generated by the cognitive agent's interaction with that environment. In this, the role of perception via the human sensory modalities in generation of the mental representations that underlie linguistic meaning is fundamental. The present discussion assesses the viability of two artificial neural network architectures, the Self Organizing Map and the Topology Preserving Network, as mechanisms for generation of suitable representations.

1. **Introduction**

Theoretical characterization of linguistic meaning has been approached in a variety of ways. Speaks (2024) provides an overview and distinguishes 'logical' approaches in the tradition of Frege, where meaning is seen as semantic interpretation of symbols in an abstract formal system, and 'foundational' approaches which focus on the mechanism of semantic interpretation; foundational approaches are subcategorized into 'use' theories such as those of Grice, and 'mentalistic' ones which relate linguistic meaning to the structure of cognition. The present discussion takes the mentalistic approach, and more specifically adopts the tradition in Western thought (Moisl 2020) ranging from Aristotle to theories of mental content in present-day linguistics and cognitive science more generally (Adams & Aizawa 2021) that the meaning of a word is its signification of a mental concept, and a mental concept is a representation of the mind-external environment causally generated by the cognitive agent's interaction with that environment. This tradition is currently maintained in attempts to 'naturalize' the mind, that is, to see the mind as an aspect of the natural world and therefore theoretically explicable in terms of the natural sciences (Morgan & Piccinini 2017; Papineau 2020). The precursors of naturalism were empiricist philosphers like like Mill (1806-73; Macleod, 2016) and scientists like von Helmholtz (1821–1894; Patton 2023) and Mach (1838–1916; Pojman 2019); von Helmholtz stressed the importance of sensory perception of and bodily interaction with the environment in generating a coherent system of mental representation whose structure mirrors that of the environment, and Mach saw human mentality as a teleological dynamical system tending to equilibrium with the environment via sensory and enactive interaction. In the present day, the tradition exists in a variety of disciplines and approaches to the study of mind and language: philosophy of mind (Churchland 2012), evolutionary epistemology (Bradie & Harms 2020), teleological epistemology (Neander 2017), evolutionary psychology (Downes 2024), embodied cognition (Anderson 2003; Barsalou 2010; Shapiro & Spaulding 2021, Clar, 2008), cognitive linguistics (Gärdenfors 2014; Geeraerts & Cuyckens 2012), and conceptual semantics (Jackendof, 2012).

Because we interface with the real-world environment using our senses, the role of perception via the human sensorimotor system in the generation of mental representations is a prominent feature of all the foregoing approaches to naturalizing cognition. The present discussion assesses the suitability of two artificial neural network architectures, the Self-Organizing Map (SOM) and the Topology Representing Network (TRN), for functional modelling of sensory systems, the aim being to identify a mechanism for generation of representations on the basis of which linguistic meaning can be implemented.

The discussion is in five main parts: the first part is this Introduction, the second motivates the selection of the two network architectures for evaluation, the third introduces relevant mathematical

concepts, the fourth describes and assesses the architectures, and the fifth concludes.

2. **Motivation**

In philosophy of mind, 'intentionality' is used to denote 'aboutness' of mental states, '*the power of minds and mental states to be about, to represent, or to stand for, things, properties and states of affairs*' (Jacob 2023) in the mind-external world. If one discounts non-physical entities like souls, as this discussion does, then the only way of explaining the manifest existence of intentionality in humans is implementation via the structure and dynamics of the physical brain. The fundamental problem in the disciplines comprising cognitive science thereby becomes understanding of how the implementation works.

The dominant paradigm for addressing this problem in the second half of the twentieth century was the Computational Theory of Mind (CTM; Rescorla 2024), whereby the architecture of cognitive functions, including language, is seen as computation over neurally implemented recursively structured mental representations. This paradigm was challenged by the philosopher John Searle in a series of publications (Cole 2024) beginning in 1980. He distinguished two sorts of intentionality, original and derived, where original intentionality is that which humans possess, and derived intentionality that which we attribute to physical mechanisms which we have reason to believe do not have original intentionality, such as thermostats and vending machines, but also the digital computers on which the theory and modelling practice of CTM are based. Searle (1980) went on to argue that *'intentionality in human beings (and animals) is a product of causal features of the brain',* and that '*any attempt literally to create intentionality artificially (strong AI) could not succeed just by designing programs but would have to duplicate the causal powers of the human brain'.*

The validity of the latter assumption is an empirical matter, but in practical terms it makes sense to use biological brain structure as a design guide wherever possible because it is the only physical system known to implement original intentionality. Each biological sensory modality has one or more cortical regions specific to itself, where the typically very high dimensional sensory activations are represented by neural activations on the cortical surface, and these representations are subsequently processed in other areas of the brain to generate cognition (Kaas 1997; Cang & Feldheim 2013; Bednar & Wilson 2016; Eickhoff et al 2018). A prominent feature is that the similarity structure among distinct sensory activation patterns is preserved in the corresponding cortical activation locations, so that identical inputs activate the same location, closely similar inputs activate closely adjacent locations, and increasingly dissimilar inputs activate proportionately distant locations (Bednar & Wilson 2016). Because of this property, sensory cortical areas are referred to as topographic maps; 'topography' is a compound of Greek *topos*, 'place', and *graphia*, 'writing' traditionally used to describe physical landscapes, metaphorically to describe physical structure in a general sense, and in neuroscience to denote spatial distribution of physical brain activation patterns in response to input.

The most-studied of the sensory topographic maps is the visual system (Brewer & Barton 2012; Bednar & Wilson 2016; Sedigh-Sarvestani et al 2021), where afferent light activates the receptor neurons in the retina, and these activations are transmitted to the lateral geniculate nucleus (LGN) and thence to the first of the regions of the visual system, referred to as V1; V1 is a topographic map of activations in the LGN, and the LGN is a topographic map of the retina. Topographic maps have also been found in the other sensory systems (Kaas 1997; Huber et al 2020), and there are indications of their existence in brain areas which process input maps (Silver & Kastner 2009; O'Rawe & Leung 2020). They have, moreover, been identified in a range of non-human species (Kaas 1997; Bednar & Wilson 2016). Topographic representation of sensory input is increasingly regarded as a fundamental mechanism in biological neural processing (Kaas 1997; Eickhoff et al 2018; O'Rawe & Leung 2020), and so, when constructing a model of how the brain represents the organism-external world, it's a good place to start.

The underlying principle of biological topographic representation is that the similarity structure of environmental input is represented in the brain as spatial distribution of neural activation in response to input: the relative locations of the physical neural activations which causally drive brain dynamics reflect the similarity structure of mind-external objects and their interactions, and are

thereby 'about' the mind-external world without involvement of a system-external interpreter, as noted by Churchland (2012). In short, topographic representation implements Searle's original intentionality.

Moisl (2021, 2022) proposed models for word and sentence meaning based on implementation of original intentionality using the autoassociative multilayer preceptron (aMLP), an artificial neural network architecture (Aggarwal 2018) shown in Figure 1.

Figure 1. An autoassociative MLP

The input and target output of an aMLP are identical, and, subsequent to training, presentation of any vector *v* in the training set results in *v* as output. The hidden layer is a representation of the input, where 'representation' is understood in its etymological sense of a re-presentation of any given form in some different form. In Figure 1, the location of the point specified by the values of the input vector in 8-dimensional vector space is re-presented as its location vector in 4-dimensional space. Where an aMLP is trained on a set V containing two or more vectors, the similarity structure among hidden layer vectors preserves that of the component vectors of V.

Given the self-imposed constraint that modelling of original intentionality should be guided by the mechanism of the biological brain, the problem with the aMLP is that the backpropagation algorithm central to its learning capability is regarded as biologically implausible (Song et al 2020; Lillicrap et al 2020). The SOM and the TRN are both more biologically plausible in that they (i) are self-organizing, and (ii) represent the similarity structures of their input domains topographically.

3. **Concepts**

Some standard mathematical concepts are introduced at the outset for convenience of exposition later on. The presentation is intuitive; formal details are in Clapham & Nicholson (2013) and further references given as appropriate.

3.1 **Space**

Colloquially, 'space' means the physical reality that humans perceive and 'dimension' is one of three possible directions in it - length, width, and height - and one of time. In mathematical usage 'space' means a set of mathematical objects with one or more operations defined on it, and 'dimension' is a measurement of the objects comprising the set; it has no necessary interpretation in terms of perceived physical reality, and no possible interpretation for dimensions greater than 4. The physical interpretation is commonly used as a metaphor to aid conceptualization of higher-dimensional mathematical spaces, but throughout this discussion it will be important to keep in mind that it *is* a metaphor, and that the two senses of the word denote ontologically different things.

Three kinds of mathematical space are referenced here, all of them defined on a subset of the *n*-fold Cartesian product of the set of real numbers $\mathbb{R}$ which generates a set C of ordered real-valued *n*-tuples.

- A vector space V interprets C as a set of vectors, where the values in each $v \in V$ are further interpreted as coordinates of a point in an *n*-dimensional coordinate system, and defines on V the operations of addition and multiplication by a scalar.
- Given a definition of 'distance', a metric space defines a distance function *d* on the points comprising V: if, for each pair of vectors v_x and $v_y \in V$, $d(v_x v_y)$ returns a non-negative real number representing the distance between points v_x and v_y, then *d* is a metric and V becomes a metric space.
- Given a set of real-valued tuples C and a collection T of open subsets of C, the pair (C,T) constitutes a topological space subject to a few set-theoretic conditions. The concept of coordinates for vector and metric spaces is discarded, and that of distance between points relative to fixed coordinates is replaced by proximity defined in terms of set union and intersection.

These spaces are covered in numerous textbooks, for example Strang (2023) on linear algebra, Ó Searcóid (2010) on metric spaces, and Munkres (2017) on topology.

3.2 **Topological manifold**

A manifold M is a distribution of points in an *n*-dimensional vector space. The shape of the distribution can be described by imposing a topology on it, thereby transforming it into a topological manifold (Lee 2011). This is done by partitioning the manifold into a set of neighborhoods such that each point x ∈ M is associated with a neighborhood N, and each N comprises x itself together with some number of other points y ∈ M within a prescribed distance δ from x: $N(x \in M) = \{y \in M \mid d(x,y) < \delta\}$, where *d* is a distance function. All the Ns are metric subspaces of M with a shared coordinate system that allows distances between points within any neighborhood to be calculated, but proximity between the Ns is defined by topological proximity rather than by distance, so that the distance between arbitrary points on the manifold without reference to the neighborhoods is undefined. Such an M is a metric topology, and where the metric is linear M's constituent neighborhoods are locally Euclidean spaces. The connection with general topological spaces is established by interpreting the neighborhoods as homeomorphic with the open subsets of general topology.

4. **Candidate replacement architectures**

Given the requirement of biological plausibility, the replacement for the aMLP should be a self-organizing artificial neural network architecture that generates structure-preserving topographic representations of input. In addition, these representations must incorporate any nonlinearities in the input because the biological brain is a nonlinear dynamical system (Breakspear 2017) on account of the nonlinearity of neural activation dynamics and extensive feedback. Any model hoping to use biological brain mechanism as a design guide must accommodate nonlinearity in some way.

Two candidate replacement architectures are considered: the Self-Organizing Map (SOM) and the Topology Preserving Network (TRN). Exemplification is via a vector space S abstracted from a small set of short alphabetic strings. Because it will be necessary to judge intuitively how well the structure of S is preserved, these strings are constructed so that their degree of similarity is obvious from direct inspection, as shown in Table 1.

abc	**def**	**jkl**	**mno**	**xyz**
ccba	eddd	kllk	moom	yxxx
babc	dded	jjkk	mnmo	yyyx
cabc	defd	jjkj	ommn	xxyy
bbbc	ffdd	kkkl	ommm	xyyy
cbbb	ffdf	llkk	nmoo	xyxx
cbac	feef	ljkj	mmmm	yxxy
	fffd	lkkj	nmnm	xyxy
	deef			yxyy
	dedf			
	edfe			
	eefe			
	ddff			

Table 1: Example string data set S

The strings in each column are constructed by randomly selecting letters from the ones at the head, and each letter is represented by a 5 x 5 bitmap, shown in Figure 2 for 'A'

Figure 2: An example letter bitmap

The bitmaps are row-wise concatenated into a 25-element vector, so that the 'A' above looks like Figure 3, where a white square = 0 and a black one = 1.

Figure 3: The bitmap in Figure 6 row-concatenated

For a four-letter string the letter vectors are concatenated to form a 100-element vector, which yields the input matrix S of 40 data points in 100-dimensional space.

4.1 **SOM**

The SOM is the obvious replacement choice both because it fulfills the above criteria and because it was designed specifically to model biological topographic organization (Kohonen 2001; Ritter et al 1992; Yin 2008; Astudillo & Oommen 2014). This section first briefly describes the SOM and then asseses is suitability with respect to the criteria.

4.1.1 **Description**

A physical SOM consists of two layers of artificial neurons or 'units' with connections between them. The arrangement of input units in physical space is undefined, but the output units are arranged on a two-dimensional grid such that the distance between all adjacent pairs of units is constant. The units comprising the first layer receives signals from an environment, and these are propagated along the connections to the second, each of whose units sums the signals it receives along the connections incident on it and is thereby activated, constituting its response to input. The degree of activation of any unit is contingent on variation in the efficiency with which input signals are propagated along its connections, and this variation is learned from the pattern of similarity relations in the input data using the training algorithm described below. The distributional pattern of all unit activations in the second layer is the network's response to the current input, and in terms of the present discussion is the representation of that input. Figure 4 is a graphical representation of a physical SOM. Only a small selection of connections between layers is shown to avoid clutter, but in reality each input unit is , connected to all output units, resulting in a dense connectivity architecture.

Figure 4: Graphical representation of physical SOM structure and connectivity

For computational simulation, a SOM's components are mathematically represented.

- Physical input signals are *n*-dimensional real-valued numerical vectors, and the set of *m* input vectors constitutes a set of points in a vector space represented computationally by a matrix M with *m* rows and *n* columns.
- The input units are represented as a vector v having the same dimension *n* as the input vectors.
- The output units are characterized mathematically as a lattice, here understood as a manifold of dimension 2 embedded in the real-number space R^n in which all pairs of adjacent points are at a constant distance from one another. For computational purposes this manifold is represented as a two-dimensional matrix O with *p* rows and *q* columns. The numerical value at $O_{i,j}$, $i = 1...p$, $j = 1...q$, represents the activation of the referenced unit.
- Each unit $O_{i,j}$ is associated with an *n*-dimensional vector *w,* its weight vector, each component of which numerically represents the efficiency of one physical connection, usually though not necessarily as a real value in the interval 0...1, between input units and the output unit $O_{i,j}$
- There are $c = p \times q$ weight vectors *w,* and these are represented computationally by a connection strength matrix W with *c* rows and *n* columns.

- The propagation of any given input signal M_i, i = 1...*m*, through the connections of a trained network to activate O is represented as the dot product of M_i and each W_j, *j* = 1...*c*. W is thereby the function that maps the input vectors to units in O.

Training proceeds as follows:

1. Initialize W with random real values in the interval 0...1.
2. Randomly select M_i, *i* = 1...*m*, and copy it to the vector *v* representing the input units..
3. Calculate the Euclidean distance *d* between *v* and each W_j, *j* = 1...*c*.

$$d = \sqrt{\sum_{k = 1...n} (v_k - W_{j,k})^2}$$

The W_j with the smallest distance from *v* best matches the input vector, and the unit in O associated with it is designated the *bmu*, 'best matching unit'.

4. Update the W_j associated with the *bmu* to make it more similar to the current input vector *v:*

$$W_{bmu} = W_{bmu} + (v - W_{bmu})$$

5. Also update the connection vectors in the neighbourhood *h* of the *bmu*, but to a decreasing degree. This is key to the operation of the SOM and is parameterized in several ways: the shape of the neighbourhood, its size, and the relationship between distance from the *bmu* and the size of the update. A frequently-used shape, used here, is square, as shown in Figure 5, where the *bmu* is represented as black, denoting the greatest degree of update, and the lightening greyscale surrounding it represents the diminishing degree of update in some researcher-defined proportion of distance from the *bmu*. Training begins with an initial neighbourhood size which diminishes as training proceeds.

Figure 5: A *bmu*, shown black, with neighbourhood on a SOM lattice

There is also a prespecified initial learning rate σ which scales the size of the update and decreases as training proceeds:

$$W_{bmu} = W_{bmu} + \sigma (v - W_{bmu})$$

6. Repeat (2) - (5) until connection weights stabilize.

When any vector $v \in M_i$ is propagated through the W of a trained network, the value generated by the dot product of *v* with each of the vectors W_j, *j* = 1...*c*, generates a pattern of activation on O, and that pattern is a representation of *v*.

4.1.2 **Assessment**

The SOM satisfies both the self-organization and nonlinearity requirements but it can fail with respect to topographic representation.

Nonlinearity

An aMLP constructs its representations via the nonlinear activation function in its hidden layer. The SOM lacks an intervening nonlinearity between layers: output activation is via matrix-vector multiplication, a linear operation. It does, however, incorporate nonlinearity by constructing a piecewise-linear approximation to any curvature in the input manifold.

Cartography provides an intuition. The spherical shape of the Earth is approximated by covering the surface with a patchwork of locally-flat regions which is projected onto a flat map. The map is distorted relative to the true shape of the Earth, but the distortion diminishes as the size of the patches decreases. Mathematically this is piecewise-linear approximation, where the collection of contiguous or intersecting locally-Euclidean point-sets of a topological manifold follows the shape of the manifold.

The SOM transforms an input vector space V into a topological manifold M via partition of the $v \in V$

into locally-linear neighborhoods using Voronoi tesselation (Aurenhammer et al 2013; Lazar et al 2022). Intuitively, a tesselation is a covering of a surface with tiles of some shape such that there are no overlaps or gaps between the tiles. Mathematically, a tesselation is a covering of a manifold surface by some number of sets which meet only on their boundaries. Construction of such sets by the SOM uses vector quantization: given a set of m real-valued vectors V, a vector quantization (VQ; Gersho & Gray 2011) of V is a collection of k subsets of V, $k < m$, centred on a set P of prototype vectors or centroids having the same dimension as those of V, where each P_i, $i = 1...k$, is associated with a subset of vectors $\{V_i\}$ such that each member of $\{V_i\}$ is closer to P_i than it is to any other P, given a distance function d. VQ thereby partitions V into k disjoint subsets yielding a manifold M, the i'th partition of which is defined as

$$M_i = \{v \in V \mid ||P_i - v|| \leq ||P_j - v||\} \text{ where } i, j = 1...k \text{ and } i <> j$$

The vectors of the original data are replaced by a set of prototypes which are representative of the data in the sense that the distribution of the prototypes approximates the data's probability density function.

Graphically, a Voronoi-tessellated region of a two-dimensional manifold looks like Figure 6 - the prototypes $p \in P$ are shown as dots, and the boundaries enclose all the vectors $v \in V$ which satisfy the above equation.

Figure 6: Voronoi cells in a magnified fragment of a manifold

Each Voronoi cell in Figure 6 is centred on a prototype. The prototypes are vectors having the same dimension as the vectors comprising M and so index locations on or near M, thereby approximating the shape of M including any curvature, and the locally linear Voronoi partitions enclosing the prototypes consequently follow the possibly-nonlinear shape of M.

With M and W represented as matrices, propagation of inputs through the connections constitutes a function $\Phi = W.M^T$, a dot product and therefore a linear operation. Every linear operation is a homomorphism, which in linear algebra (Strang 2023**;** Butterfield et al 2016, 'Homomorphism') is a linear mapping between two vector spaces and in metric space terms a linear mapping between two spaces M1 and M2 which preserves the distance structure of M1, that is, the inter-point distances between all $v_i, v_j \in$ M1 in M2, given a distance function d. The function Φ is therefore a homomorphic mapping of a piecewise-linear approximation to the possibly-nonlinear shape of M to the SOM's output space, and in this way the SOM can incorporate nonlinearity in its representation of input.

Topographic representation

The SOM interprets manifold structure preservation topologically as homeomorphism (Munkres 2017, 105) - a function f:A > B that maps one topological space A to another B such that f is (i) bijective, that is, one-to-one and onto, (ii) continuous, and (iii) invertible as a continuous function f^{-1}: B > A. Intuitively, given any two topological spaces A and B, 'homeomorphism' denotes their indistinguishability: f:A >B preserves the topological properties of A in B, which include nearness and connectedness but not, crucially, metric distance and therefore not manifold shape - a circle can be homeomorphically mapped to a square and vice versa, and so the two shapes are topologically indistinguishable. Note, incidentally, that homeomorphism (with and 'e') has nothing to do with the homomorphism mentioned earlier.

The SOM's topological spaces are input and output manifolds M and O respectively, and the learned function f:A > B is a mapping from an n-dimensional M to a 2-dimensional O. Therein lies a problem. The prespecification of dimension-2 for O raises the possibility of a dimensional mismatch with M. Brouwer's invariance of domain theorem (Munkres 2017, 383-4) states that, given an open subset U of the Euclidean space R^n for any integer n, if f:U > R^n is a continuous injective map, then V = f(U) is also an open subset of R^n. A corollary is invariance of dimension (Lee 2010, 40): if, for two integers m and n, $m = n$, then U and V are homeomorphic, but otherwise not. In other words, for a continuous injective map between two Euclidean spaces to be homeomorphic, the spaces must have the same dimension. This applies to the SOM because its training algorithm transforms the input data into a topological manifold - a collection of locally-Euclidean open

subsets whose union is an open subset with dimension equal to that of the subsets (Munkres 2017, 213-15; Lee 2010, 41). The implication is that, for the SOM's lattice accurately to represent the topology of the input manifold topographically, the input manifold must have dimension-2 because that is the predefined dimension of the lattice. But sensory input to the brain is very high-dimensional, and the intuition is that the level of dimensionality mismatch involved is bound to invalidate the SOM as a model for neural topographic representation of real-world input.

Intuition can be unreliable, and this one is potentially mitigated by the manifold hypothesis underlying data dimensionality reduction methods (Ghojogh et al 2023), which says that data derived from the natural world is typically redundant - that data of dimension *n* can be restated using *m* dimensions, $m < n$, without significant loss, but there is a lower limit. That limit is the intrinsic dimension of the data. Intrinsic dimension is the minimum number of variables *m* required to describe what the given data describes using *n* variables. More formally, the location of any point *p* in a vector space V is specified by a vector whose *n* elements are the coordinates of *p* relative to the *n* axes of V, for some integer *n*. The minimum *n* required to specify the location of points comprising a geometrical object embedded in a space is its intrinsic dimension (Camastra & Staiano 2016): a line, for example, has intrinsic dimension 1, its length, a plane has intrinsic dimension 2, length and width, and so on to higher dimensions. An object of intrinsic dimension *n* can be embedded in a higher-dimensional space *n+k*, for $k = 1, 2...$, in that it can be described by *(n+k)*-dimensional vectors with reference to *n+k* axes, but its intrinsic dimension remains the same - a line is only ever a 1-dimensional object irrespective of embedding dimension. One could argue that the SOM's homeomorphic mapping is contingent on the intrinsic rather than on the observed input data dimension - if the intrinsic dimension of real-world input were to be identical or at least near-identical to that of the SOM's output lattice, the SOM would remain viable as a model, but current indications are not encouraging. Pope et al (2021) have investigated the intrinsic dimension of several large image data sets and concluded that these do *'have very low intrinsic dimension relative to the high number of pixels in the images*', but their intrinsic dimension lies between 26 and 43 - still far from the SOM's 2.

To exemplify, a SOM with a 12 x 12 lattice was trained on the above string data S. Figure 7 shows the superimposed *bmu* activations in response to all 40 strings after training. For visual clarity the lattice is shown as a mesh plot with *bmu* locations at the line intersections.

Figure 7. Lattice *bmu* activations for all 40 input vectors of S shown simultaneously

Visual inspection reveals a spatial arrangement that resembles the neighborhood structure of the input manifold shown in Figure 8, but that structure is distorted. The eye picks out clusters of similar strings, but the placement of the 'def' series appears to contravene the definition of a Voronoi cell as a subset whose members are closer to their prototype vector than to prototype vectors of any other cells.

Figure 8. Neighborhood structure of the data set S

This visual impression is confirmed in Figure 9 by cluster analysis of the activation map coordinates in Figure 7.

ccba	11	12
babc	11	10
cabc	12	10
bbbc	12	6
cbbb	12	8
cbac	10	8
eddd	12	4
dded	10	6
defd	6	3

ffdd	8	1	
ffdf	10	1	
...	...	...	
xyxy	2	12	
yxyy	2	8	
bmu coordinates			Lattice clustering

Figure 9. Single link clustering of *bmu* coordinates on the lattice in Figure 7

The grouping of strings in Figure 9 resembles that of the Voronoi partition in Figure 8, but 'def' series strings are spread across three different clusters. Many copies of the SOM variously parameterized were tried, always with analogous results: points which are topologically close in M are usually but not always represented as close on O, and just because points are close on O does not imply that they are close on M.

This problem is intrinsic to the SOM. The relevant literature standardly claims that the topology of an input manifold M is represented topographically by the relative distances between unit activations on O, but it also typically qualifies this by noting that the SOM preserves input topology topographically 'as well as possible' or some similar formulation, and provides criteria for assessing the degree to which topology has been preserved (Pölzlbauer 2004; Hamel 2016).

Given that the SOM lattice is intended to be physically interpreted as the 2-dimensional surface of the biological cortex, the conclusion is that the SOM's topographic representation of input structure is unreliable.

4.2 **The Topology Representing Network (TRN)**

Martinetz & Schulten (1994) recognized the SOM's dimension mismatch problem and proposed the Topology Preserving Network (TRN) to resolve it. The TRN learns to represent the input topology as a graph whose nodes are interpretable as the units / neurons of a physical network, whose links are the physical connections between them, and the activation patterns of which represent input topology topographically. Crucially, the TRN does not reduce input manifold dimension and so does not suffer from the SOM's dimension-mismatch problem.

4.2.1 **Description**

The TRN differs from the SOM in that its output units have no predefined spatial structure . A physical TRN prior to training looks like Figure 10.

Figure 10: A physical TRN

Like the SOM, external signals are loaded into the input units and transmitted to the output units via the connections. Each output unit is connected to all the inputs, though as before only a few of these are depicted. The output units are shown as a sequence, but there is no implication of any specific structure. Also like the SOM, training proceeds by repeated presentation of vectors selected randomly from an input set, but adjustment in reaction to a presentation differs, as described below.

Mathematically:

- Physical inputs are represented as *n*-dimensional real-valued vectors, and the set of *m* input vectors constitutes a vector space represented by an *m* x *n* matrix V.
- The *p* output units are represented by a *p*-component vector *o*.
- The connections are represented by a real-valued *p* x *n* matrix W each row W_j of which, *j* = 1...*p*, represents the connections from output unit o_j to the *n* input units.
- The transformation of a given input signal to a pattern of output unit activations is

represented as the dot product of the corresponding input vector V_i and the connection vector W_j associated with each of the p output units, $j = 1...p$, so that the output vector o generated by any given V_i is

$$o_{j,\ j = 1...p} = V_i \cdot W_j$$

The TRN is trained in two conceptually separate steps. At the first step, V is Voronoi-partitioned such that the connection vector for any output unit o_j, $j = 1...p$, constitutes the prototype vector of a Voronoi cell, which transforms V into a topological manifold M and locates the output units in M via the prototype vector coordinates, like the SOM. Training proceeds in a sequence of iterations, making incremental changes to the connection strengths at each iteration so as to optimize a criterion. The criterion in this case is the absolute global difference between the input vectors V and the connections W, that is, to minimize the error E:

$$E = \sum_{i=1...m,\ j = 1...p} ||V_i - W_j||$$

Two parameters are prespecified: (i) a learning rate r which scales the magnitude of the connection updates, starting with a large value and gradually decreasing as training proceeds, and (ii) a neighborhood λ which, for any output unit o_j is the number of units near o_j; 'near' is defined in what follows. Like the learning rate, the neighborhood starts large and decreases incrementally.

At each iteration an input vector $v = V_i$, $i = 1...m$, is randomly selected, the absolute difference between it and each of the connection vectors W is calculated, and the result is stored in a difference table T:

$$T_j = ||v - W_j||,\ j = 1...p$$

T is sorted in ascending order of difference magnitudes and the corresponding connections are updated in proportion to their position in the table: the connection vector with the smallest difference from v is updated most, the connection vector with the second-smallest difference from v is updated second-most, and so on up to the current neighborhood limit λ. 'Updating' means adjusting the connection vector W_k indexed by position T_j, $j = 1...p$, so as to make it more similar to the current input vector v:

$$W_k^{new} = W_k^{old} + r(e^{-j/\lambda}(v - W_k^{old}))$$

where e is Euler's number and j is the index of W_k in the sorted table T. The value of the $(e^{-j/\lambda}(v - W_k^{old})$ term decreases with increasing j so that the update effects incrementally diminish.

The result of training is a connection matrix W whose row vectors specify locations in the input manifold M, and, since the vectors in W have the same dimension as those in M, these locations are the prototype vectors of Voronoi cells.

At the second step, a graph representation of the adjacency of the Voronoi cells is abstracted from the tesselation. This representation is a Delaunay graph D (Aurenhammer & Klein 2013), which is constructed by connecting all pairs of prototype vectors W whose Voronoi cells share a boundary. This is shown for the 2-dimensional case in Figure 11.

Figure 11: Delaunay graph of a Voronoi tesselation, from Martinetz & Schulten (1994)

D is a set of graph vertices $D = \{W_1...W_p\}$ in which any two vertices are connected by an edge if their Voronoi cells are adjacent.

Though construction of the Delaunay graph is conceptually separate from tthe Voronoi partition, the TRN inserts it as a step in the training algorithm: once the difference table T has been sorted in ascending order, a link is created between the unit whose connection vector that is nearest to the current input v and the unit whose connection vector is second-most similar to v if one does not already exist. The effect is incrementally to create a graph structure of nearest-neigbour links as training proceeds. This incorporation of graph creation into the training algorithm introduces a complication in that links created at any stage of training may become obsolete as the prototype vectors of the Voronoi cells converge from initial randomly-assigned locations in M to their final

locations. The solution is a predefined threshold whereby, at each training step, the links 'age' - the age of a newly-created link, or of an existing link which the above procedure reiterates, is set to 0; in all other cases the age is incremented and any links whose age reaches the threshold are removed.

4.2.2 **Assessment**

Nonlinearity

Like the SOM, the TRN Voronoi-tesselates the input space, so the same argument as for the SOM applies.

Topographic representation

Because the TRN does not prespecify the dimension of its output space, a topographic representation of input structure analogous to that of the SOM appears not to be possible because an arbitrarily large number *n* of output units is permitted, and as *n* grows so does the dimension of the mathematical output space; once *n* exceeds 4 the space has no physical interpretation. It therefore appears that the TRN is unsuitable in general for modelling of topographic representation.

There is, however, an alternative interpretation based on the Delaunay graph. Because the graph arcs join the centroids of spatially contiguous neighborhoods they represent not just neighborhood connectivity but also the pattern of neighborhood spatial distribution. The Delaunay graph generated by a TRN is thereby a model of the shape of the input manifold. Physically interpreted, it represents the activation of some specified number of artificial or biological neurons in response to environmental input, the spatial activation patterning of which mirrors the topological structure of that input - in other words, it models topographic organization.

To exemplify, a TRN with 18 output units was trained using S. Figure 12 shows the resulting Delaunay graph with numbered output unit nodes after training, and Table 2 the corresponding connectivity matrix with arc connection values highlighted.

Figure 12. Delaunay graph generated by a TRN with 18 output units, with nodes labelled by the strings included in the corresponding Voronoi neighborhoods

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
1	0	0	0	0	0	0	0	0	0	**0.29**	0	0	0	0	0	0	**0.62**	0
2	0	0	0	0	0	0	0	0	0	0	0	0	0	**0.87**	0	0	0	**0.91**
3	0	0	0	0	0	0	**0.01**	0	0	0	0	0	0	0	0	0	**0.96**	0
4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	**0.89**	0	0
5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	**0.97**	0	0
6	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	**0.73**	0	0
7	0	0	**0.99**	0	0	0	0	0	0	0	0	0	0	0	0	0	**0.85**	0
8	0	0	0	0	0	0	0	0	0	0	0	0	0	**0.84**	0	0	0	0
9	0	0	0	0	0	0	0	0	0	0	0	0	0	0	**0.65**	0	0	0

10	**0.82**	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
11	0	0	0	0	0	0	0	**0.76**	0	0	0	**0.94**	0	0	0	0	0	0
12	0	0	0	0	0	0	0	0	0	0	**0.92**	0	0	0	0	0	0	0
13	0	0	**0**	0	0	0	0	0	0	0	0	0	0	**0.22**	0	0	0	**0.81**
14	0	**0.95**	0	0	0	0	0	**0.69**	0	0	0	0	0	0	0	0	0	0
15	**0.54**	0	0	0	0	0	0	0	**0.74**	0	0	0	0	0	0	0	0	0
16	0	0	0	0	0	**0.9**	0	0	0	0	0	0	0	0	0	0	0	0
17	0	0	**0.98**	0	0	0	**0.75**	0	0	0	0	0	0	0	0	0	0	0
18	0	**0.35**	0	0	0	0	0	0	0	0	0	0	**0.83**	0	0	0	0	0

Table 2. Connectivity matrix of the graph in Figure 12

There are four disconnected subgraphs with nodes labelled by the strings which map to them, that is, by the constituents of the associated Voronoi neighborhoods. Comparison of the graph with the neighborhood structure of M in Figure 8 shows them to correspond in terms of string grouping, and so the connectivity structure can be said to have preserved the input topology of S. Because a Voronoi neighborhood contains all points of a manifold closer to their centroid than to the centroid of any other neighborhood, and because Voronoi neigborhoods with Delaunay links are contiguous, topologically close input vectors will always be close in the similarity structure of the corresponding output patterns; Martinetz & Schulten (1994) prove this for manifolds of sufficient density to describe the manifold well. Note that the spatial distribution shown in Figure 12 is an artefact of the plotting software.Other distribution patterns having the same connectivity are possible; the point is that connected nodes are guaranteed to be spatially close.

In graph theory every finite graph can be embedded in three-dimensional Euclidean space (Archdeacon 1996; Cohen et al 1997; Boutin 2005), where embedding is understood as a representation of an object O in a space S1 in another space S2 such that the properties - in the present case connectivity - of O_{S1} are preserved in O_{S2}. When the input and output spaces of a TRN are finite, this applies to the Delaunay graph which it generates. Every physical implementation of a TRN is necessarily finite, and so the mathematical finite three-dimensional embedding space is directly interpretable in terms of the three spatial dimensions of physical reality. In the present context the nodes of Figure 12 are interpreted as the units of an artificial or biological network and the arcs are connections among them as specified in Table 2; only the first two rows of Table 2 are shown in Figure 13.

Figure 13. Physical interpretation of a TRN implementing the graph structure of Figure 12

Activation of any particular output unit o_i, i = 1...18, is an additive combination of the activation generated by an input vector via the receptive field of o_i and the activations of output units to which o_i is linked by lateral connections.

$$o_i = (M_j.C_i) + (\sum_k(o_k L_k))$$

where:

- M_j is the input manifold indexed by the *j*'th vector for j = 1...40.
- C_i is the input-to-output connection matrix indexed by the *i*'th vector of connection values

representing the physical receptive field of o_i.

- $M_j.C_i$ is the dot product.
- $\sum_k(o_k L_k)$ is the sum, for all k output units to which o_i is laterally connected, of the activation of o_k multiplied by the lateral connection strength L_k, where L is the lateral connection matrix shown in Table 2.

5. **Conclusion**

The aim of this discussion has been to identify an alternative to the autoassociative multilayer perceptron for implementation of original intentionality and thereby of linguistic meaning, motivated by Searle's conviction that any viable mechanism for this purpose will require '*the causal powers of the human brain'.* This implies an artificial neural network, for which three criteria were specified: that it must be self-organizing, that it must accommodate any nonlinearity present in the input data, and that the representations it generates must be interpretable as implementations of biological topographic representation of sensory input. Both the SOM and the TRN satisfy the first two criteria, but the SOM fails with respect to the third in that any mismatch of input space dimension with the SOM's predefined 2-dimensional output space potentially distorts representation of the data manifold and thereby the reliability of the topographic representation on the output lattice. The TRN overcomes the dimension-mismatch problem by eliminating the output lattice and instead constructing a Delaunay graph which, assuming a sufficiently dense input manifold point distribution, is proven by its designers to represent the input topology optimally for a network of a predefined size, and which is interpretable as topographic organization. The conclusion is that the TRN is preferable to the SOM as a mechanism for generation of biologically-plausible, input topology preserving topographic maps usable for implementation of original intentionality and thereby of intrinsic linguistic meaning.

Naturalism in cognitive science, including linguistics, sees the mind as an aspect of the natural world and therefore theoretically explicable in terms of the natural sciences, as noted at the outset. The language of the natural sciences is mathematics, and the *Journal of Quantitative Linguistics* is committed to theoretical characterization of natural language in mathematical and statistical terms. The present discussion provides an example of how this commitment can be extended to understanding of linguistic meaning.

References

- Adams, F., Aizawa, K. (2021). Causal theories of mental content, *Stanford Encyclopedia of Philosophy*, *The Stanford Encyclopedia of Philosophy* (Fall 2021 Edition), Edward N. Zalta (ed.), URL = <https://plato.stanford.edu/archives/fall2021/entries/content-causal/>
- Aggarwal, C. (2018) Neural networks and deep learning, Springer
- Anderson, M. (2003). Embodied cognition: a field guide. Artificial Intelligence. 149, 91–130
- Archdeacon, D. (1996) Topological Graph Theory. A survey, Congressus Numerantium 115, 5-54
- Astudillo, C., Oommen, B. (2014) Topology-oriented self-organizing maps: A survey, Pattern Analysis and Applications 17, 223-48
- Aurenhammer, F., Klein, R., Lee, D. (2013) Voronoi diagrams and Delaunay triangulations, World Scientific
- Barsalou, L. (2010). Grounded cognition: past, present, and future. Topics in Cognitive Science, 2, 716-724
- Bednar, J., Wilson, S. (2016) Cortical maps, Neuroscientist 22, 604-17
- Boutin, D. (2005) Isometrically embedded graphs, Ars Combinatoria 77

- Bradie, M., Harms, W. (2020). Evolutionary Epistemology. The Stanford Encyclopedia of Philosophy (Spring 2020 Edition), ed. E. Zalta, URL = <https://plato.stanford.edu/archives/spr2020/entries/epistemology-evolutionary/>
- Breakspear, M. (2017) Dynamic models of large-scale brain activity, Nature Neuroscience 20, 340-52
- Brewer, A., Barton, B (2012) Visual Field Map Organization in Human Visual Cortex, in Visual Cortex- Current Status and Perspectives, ed. S. Molotchnikoff & J. Rouat, chapter 2, InTech
- Butterfield, A., Ekembe Ngondi, G., Kerr, A. (2016) A Dictionary of Computer Science, 7th ed., Oxford University Press
- Camastra, F., Staiano, A. (2015) Intrinsic dimension estimation: Advances and open problems, Information Sciences 328, 26-41
- Cang, J., Feldheim, D. (2013) Developmental mechanisms of topographic map formation and alignment., Annual Review of Neuroscience 36, 51-77
- Churchland, P. (2012) Plato's camera. How the physical brain captures a landscape of abstract universals, MIT Press
- Clapham, C.,Nicholson, J. (2013) The Concise Oxford Dictionary of Mathematics, 6th ed., Oxford University Press
- Clark, A. (2008) Supersizing the mind: embodiment, action, and cognitive extension, Oxford University Press
- Cohen, R, Eades, P., Tao, L., Ruskey, F. (1997) Three-dimensional graph drawing, Algorithmica 17, 199-208
- Cole, D. (2024). The Chinese Room Argument. The Stanford Encyclopedia of Philosophy (Winter 2024 Edition), Edward N. Zalta & Uri Nodelman (eds.), URL = <https://plato.stanford.edu/archives/win2024/entries/chinese-room/>
- Downes, S. (2024). Evolutionary Psychology. The Stanford Encyclopedia of Philosophy (Spring 2024 Edition), Edward N. Zalta & Uri Nodelman (eds.), URL = <https://plato.stanford.edu/archives/spr2024/entries/evolutionary-psychology/>
- Eickhoff, S., Constable, R., Yeo, B. (2018) Topographic organization of the cerebral cortex and brain cartography, Neuroimage 170, 332-47
- Gärdenfors, P. (2014) Geometry of meaning. Semantics based on conceptual spaces. MIT Press
- Geeraerts, D., Cuyckens, H. (2012). Introducing cognitive linguistics. Oxford University Press
- Gersho, A., Gray, R. (2011) Vector quantization and signal compression, Springer
- Ghojogh, B., Crowley, M., Karray, F., Ghodsi, A. (2023) Elements of Dimensionality Reduction and Manifold Learning, Springer
- Hamel, L. (2016) SOM Quality Measures: An Efficient Statistical Approach, in Advances in Self-Organizing Maps and Learning Vector Quantization, Merényi, E., M., Mendenhall, M., O'Driscoll, P. (eds.), Springer
- Jackendoff, R. (2012). *A User's Guide to Thought and Meaning*. Oxford University Press.
- Jacob, P. (2023). Intentionality. The Stanford Encyclopedia of Philosophy (Spring 2023 Edition), Edward N. Zalta & Uri Nodelman (eds.), URL = <https://plato.stanford.edu/archives/spr2023/entries/intentionality/>

- Kaas, J. (1997) Topographic Maps are Fundamental to Sensory Processing, Brain Research Bulletin 44, 107–112
- Kohonen, T. (2001), Self-Organizing Maps, 3rd ed., Springer
- Lazar, E., Lu, J., Rycroft, C. (2022) Voronoi cell analysis: The shapes of particle systems, American Journal of Physics 90, 469-80
- Lee, J. (2010) Introduction to Topological Manifolds, 2nd ed. Springer
- Lillicrap, T., Santoro, A., Marris, L., Akerman, C., & Hinton, G. (2020) Backpropagation and the brain, Nature Reviews Neuroscience 21, 335–346
- Macleod, C. (2016). John Stuart Mill. The Stanford Encyclopedia of Philosophy (Summer 2020 Edition), Edward N. Zalta (ed.), URL = <https://plato.stanford.edu/archives/sum2020/entries/mill/>
- Martinetz, T., Schulten, K. (1994) Topology representing networks, Neural Networks 7, 507-522
- Moisl, H. (2020) Intrinsic intentionality and linguistic meaning: An historical outline. In: Words and Numbers. In Memory of Peter Grzybek (1957-2019), ed. E.Kelih & R. Köhler, RAM-Verlag, 148-166
- Moisl, H. (2021) Implementation of intrinsic natural language lexical intentionality, Academia Letters 2021
- Moisl, H. (2022) Dynamical systems implementation of intrinsic sentence meaning, Minds and Machines 32 DOI: 10.1007/s11023-022-09590-1
- Morgan, A., Piccinini, G. (2017). Towards a cognitive neuroscience of intentionality. Minds and Machines, 28, 119-139
- Munkres, J. (2017) Topology, 2nd ed., Pearson
- Neander, K. (2017). A mark of the mental: In defense of informational teleosemantics. MIT Press
- O'Rawe, J., Leung, H. (2020) Topographic mapping as a basic principle of functional organization for visual and prefrontal functional connectivity, eNeuro 27, https://doi.org/10.1523/ENEURO.0532-19.2019
- Ó Searcóid, M. (2010) Metric Spaces, Springer
- Papineau, D. (2020). Naturalism. The Stanford Encyclopedia of Philosophy (Summer 2020 Edition), ed. E. Zalta, URL = <https://plato.stanford.edu/archives/sum2020/entries/naturalism/>
- Patton, L. (2023). Hermann von Helmholtz. The Stanford Encyclopedia of Philosophy (Fall 2024 Edition), Edward N. Zalta & Uri Nodelman (eds.), URL = <https://plato.stanford.edu/archives/fall2024/entries/hermann-helmholtz/>
- Pölzlbauer, G. (2004) Survey and comparison of quality measures for self-organizing maps, Proceedings of the Fifth Workshop on Data Analysis (WDA'04), 67--82
- Pojman, P. (2019). Ernst Mach. The Stanford Encyclopedia of Philosophy (Winter 2023 Edition), Edward N. Zalta & Uri Nodelman (eds.), URL = <https://plato.stanford.edu/archives/win2023/entries/ernst-mach/>
- Pope, P., Zhu, C., Abdelkader, A., Goldblum, M., Goldstein, T. (2021) The intrinsic dimension of images and its impact on learning, International Conference on Learning Representations 2021, https://doi.org/10.48550/arXiv.2104.08894
- Rescorla, M. (2024) The Computational Theory of Mind, *The Stanford Encyclopedia of Philosophy* (Winter 2024 Edition), Edward N. Zalta & Uri Nodelman (eds.), URL =

<https://plato.stanford.edu/archives/win2024/entries/computational-mind/>

- Ritter H, Martinetz T, Schulten K (1992) *Neural Computation and Self-organising Maps: An Introduction*. Addison-Wesley
- Searle, J. (1980). Minds, Brains and Programs. Behavioral and Brain Sciences 3, 417–57
- Sedigh-Sarvestani, M., Lee, K., Jaepel, J., Satterfield, R., Shultz, N., Fitzpatrick, D. (2021) Sinusoidal transformation of the visual field is the basis for periodic maps in areaV2, Neuron 109, 4068-4079
- Silver, M., Kastner, S. (2009) Topographic maps in human frontal and parietal cortex, *Trends Cognitive Science* 13, 488–495
- Song , Y. Lukasiewicz, T., Xu, Z. Bogacz3, R. (2020) Can the Brain Do Backpropagation? — Exact Implementation of Backpropagation in Predictive Coding Networks, 34th Conference on Neural Information Processing Systems (NeurIPS 2020), Vancouver, Canada
- Speaks, J. (2024). Theories of meaning. *The Stanford Encyclopedia of Philosophy* (Winter 2024 Edition), Edward N. Zalta & Uri Nodelman (eds.), URL = <https://plato.stanford.edu/archives/win2024/entries/meaning/>
- Strang, G. (2023) Introduction to linear algebra, 6th ed., Wellesley-Cambridge Press
- Shapiro, L., Spaulding, S.(2021) Embodied Cognition. The Stanford Encyclopedia of Philosophy (Fall 2024 Edition), Edward N. Zalta & Uri Nodelman (eds.), URL = <https://plato.stanford.edu/archives/fall2024/entries/embodied-cognition/>
- Yin, H. (2008) The Self-Organizing Maps: Background, Theories, Extensions and Applications, Studies in Computational Intelligence 115, 715–762