Springer, Heidelberg, 2000, pp. 348-362.

Self-Organizing Maps in Symbol Processing

Timo Honkela

Media Lab, University of Art and Design,
Hämeentie 135 C, FIN-00560 Helsinki, Finland
Timo.Honkela@uiah.fi
http://www.mlab.uiah.fi/~timo/

Abstract. A symbol as such is disassociated from the world. In addition, as a discrete entity a symbol does not mirror all the details of the portion of the world that it is meant to refer to. Humans establish the association between the symbols and the referenced domain — the words and the world — through a long learning process in a community. This paper studies how Kohonen self-organizing maps can be used for modeling the learning process needed in order to create a conceptual space based on a relevant context with which the symbols are associated. The categories that emerge in the self-organizing process and their implicitness are considered as well as the possibilities to model contextuality, subjectivity and intersubjectivity of interpretation.

1 Introduction

Models of natural language may test the background assumptions of the developers, or, at least, reflect them. In the predominant approaches among computerized models of language, the linguistic categories and rules are predetermined and coded by hand explicitly as symbolic representations. The field of connectionist natural language processing, based on the use of artificial neural networks, may be characterized to take an opposite stand. The critical view on symbolic representations is based on the idea that the symbolic and discrete nature of written expressions in natural language does not imply that symbolic descriptions of linguistic phenomena are sufficient as such. This view appears to be relevant especially when semantic and pragmatic issues are considered. A traditional, logic-based analysis studies examples like the ones given below (from [35]). The emphasis lies in phenomena that are suitable to be explained in the framework of predicate logic, e.g., propositional forms, connectives, quantifiers, truth values, presuppositions, and logical ambiguity.

- *"Each one of Mozart's works is a masterpiece."*
- *"If butter is heated, it melts."*

However, if one considers the sentences and conversations given below, it should be apparent that there is need for formalisms and tools that enable modeling, for instance, of adaptation, vagueness, contextuality, subjectivity of interpretation, and the relationship between discrete symbols and continuous spaces in the domain under consideration.

- *"Please, show me some pictures with beautiful Finnish lake sceneries."*
- *"Do you see that small woman there?"*
  "Actually, I don't consider her small while people in her country are usually much shorter than here."

A radically connectionist natural language processing approach is based on the following assumptions. The ability to understand natural language utterances can be learned via examples. The categories necessary in the interpretation emerge in the self-organizing learning processes and they may be implicit rather than explicit as will be shown later in this article. An implicit category can be used during interpretation even if it is not named. The processing mechanisms are mainly statistical rather than rule-like. The process of symbol grounding, i.e., associating symbols with continuous multi-dimensional spaces and dynamic processes as well as the assumptions outlined above are discussed in this paper. The methodological basis is Kohonen's self-organizing map algorithm.

2 Self-Organizing Map

The basic self-organizing map (SOM) [22, 24] can be visualized as a sheet-like neural-network array, the cells (nodes, units) of which become specifically tuned to various input signal patterns or classes of patterns in an orderly fashion. The learning process is competitive and unsupervised, meaning that no teacher is needed to define for an input the correct output, i.e., the cell into which the input is mapped. The locations of the responses in the map tend to become ordered in the learning process as if some meaningful nonlinear coordinate system for the different input features were being created over the network [24].

2.1 Self-Organizing Map Algorithm

Assume that some sample data sets (such as in Table 1) have to be mapped onto a 2-dimensional array. The set of input samples is described by a real vector $\mathbf{x}(t) \in R^n$ where t is the index of the sample, or the discrete-time coordinate. Each node i in the map contains a model vector $\mathbf{m}_i(t) \in R^n$, which has the same number of elements as the input vector $\mathbf{x}(t)$.

The initial values of the components of the model vector, $\mathbf{m}_i(t)$, may even be selected at random. In practical applications, however, the model vectors are more profitably initialized in some orderly fashion, e.g., along a two-dimensional subspace spanned by the two principal eigenvectors of the input data vectors [24].

Any input item is thought to be mapped into the location, the $\mathbf{m}_i(t)$ of which matches best with $\mathbf{x}(t)$ in some metric (e.g. Euclidean). The self-organizing algorithm creates the ordered mapping as a repetition of the following basic tasks:

1. An input vector $\mathbf{x}(t)$ is compared with all the model vectors $\mathbf{m}_i(t)$. The best-matching unit (node) on the map, i.e., the node where the model vector is

250	235	215	antique white
165	042	42	brown
222	184	135	burlywood
210	105	30	chocolate
255	127	80	coral
184	134	11	dark goldenrod
189	183	107	dark khaki
255	140	0	dark orange
233	150	122	dark salmon
...	...	...	...

Table 1. Three-dimensional input data in which each sample vector x consists of the red-green-blue values of the color shown in the rightmost column.

most similar to the input vector in some metric is identified. This best-matching unit is often called the winner.

2. The model vectors of the winner and a number of its neighboring nodes in the array are changed towards the input vector according to the learning principle specified below.

The basic idea in the SOM learning process is that, for each sample input vector $\mathbf{x}(t)$, the winner and the nodes in its neighborhood are changed closer to $\mathbf{x}(t)$ in the input data space. During the learning process, individual changes may be contradictory, but the net outcome in the process is that ordered values for the $\mathbf{m}_i(t)$ emerge over the array. If the number of available input samples is restricted, the samples must be presented reiteratively to the SOM algorithm.

Adaptation of the model vectors in the learning process takes place according to the following equations:

$\mathbf{m}_i(t+1) = \mathbf{m}_i(t) + \alpha(t)[\mathbf{x}(t) - \mathbf{m}_i(t)]$ for each $i \in N_c(t)$,

$\mathbf{m}_i(t+1) = \mathbf{m}_i(t)$ otherwise,

where t is the discrete-time index of the variables, the factor $\alpha(t) \in [0,1]$ is a scalar that defines the relative size of the learning step, and $N_c(t)$ specifies the *neighborhood* around the winner in the map array.

At the beginning of the learning process the radius of the neighborhood is fairly large, but it is made to shrink during learning. This ensures that the global order is obtained already at the beginning, whereas towards the end, as the radius gets smaller, the local corrections of the model vectors in the map will be more specific. The factor $\alpha(t)$ also decreases during learning. The resulting map is shown in Figure 1. The experiment was conducted using SOM_PAK software [25].

Perhaps the most typical notion of the SOM is to consider it as an artificial neural network model of the brain, especially of the experimentally found ordered "maps" in the cortex. There exists a lot of neurophysiological evidence to support the idea that the SOM captures some of the fundamental processing principles of the brain [27]. Other early artificial-neural-network models of self-organization have been presented, e.g., in [1], [3], and [54].

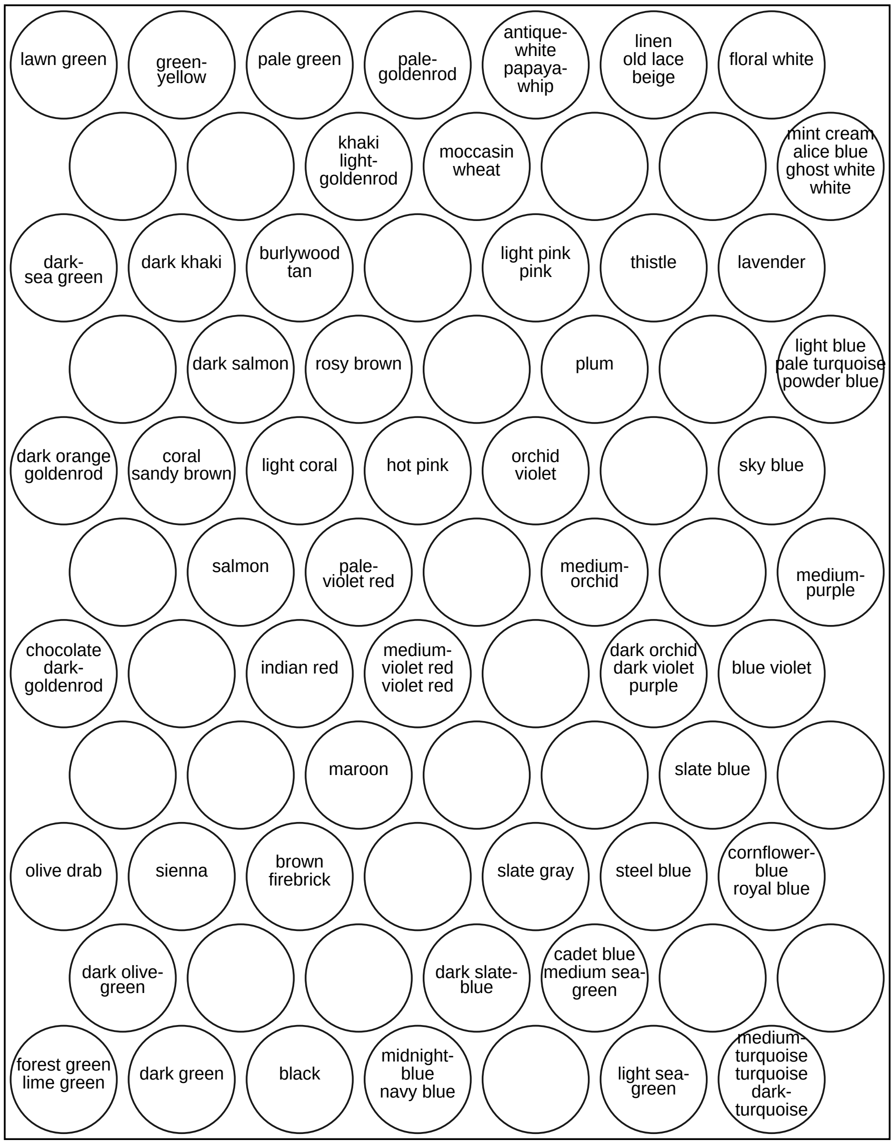

Fig. 1. A map of colors based on their red-green-blue values. The color symbols in the rightmost column of Table 1 as used in labeling the map. The best matching unit is searched for each input sample and that node is labeled accordingly.

The SOM can also be viewed as a model of unsupervised machine learning, and as an adaptive knowledge representation scheme. The traditional knowledge representation formalisms – semantic networks, frame systems, predicate logic, to provide some examples – are static and the reference relations of the elements are determined by a human. Moreover, those formalisms are based on the tacit assumption that the relationship between natural language and world is one-to-one: the world consists of objects and the relationships between the objects, and these objects and relationships have straightforward correspondence to the elements of language. An alternative point of view is that the pattern recognition process must be taken into account: expressions of languages refer to patterns and distributions of patterns in the concrete perceptual domain and often also in the abstract domain.

One does not need to question the existence of the world in order to be critical towards the notion of entities or objects as a basis for epistemological considerations. Both anticipation and context influence the perception and the naming process. This view is adopted, e.g., in constructivism. One early constructivist, Heinz von Foerster, has stated that objects and events are not primitive experiences but they are representations of relations. The construction of these relations is subjective, constrained by anatomical and cultural factors. The postulate of an external (objective) reality gives way to a reality that is determined by modes of internal computations [7]. This relativity or subjectivity of interpretation of symbols does not lead, however, into arbitrariness while the language users can refine their interpretation models closer to each other through communication. The use of the SOM in modeling such a learning process is considered in [12].

3 SOM-based Symbol Processing

The self-organizing map can be used in several ways when symbol processing is considered. One can create a map of symbols by associating each label with a numerical vector and finding corresponding best-matching location on the map. Another mean is to encode each symbol with, e.g., a unique random vector and use this coding as the basis in the learning process [47]. The order of the map is based on presenting the encoded word with its context during the learning. The context can, e.g., be textual [13, 47], or numerical measurements and representations [11]. The latter can originate from a visual source. An overview of connectionist, statistical and symbolic approaches in natural language processing and an introduction to several articles is given in [55].

3.1 Maps of Words

Contextual information has widely been used in statistical analysis of natural language corpora. Charniak [4] presents the following scheme for grouping or clustering words into classes that reflect the commonality of some property.

1. Define the properties that are taken into account and can be given a numerical value.
2. Create a vector of length n with n numerical values for each item to be classified.
3. Cluster the points that are near each other in the n-dimensional space.

The open questions are: what are the properties used in the vector, the distance metric used to decide whether two points are close to each other, and the algorithm used in clustering. The SOM does both vector quantization and clustering at the same time. Moreover, it produces a topologically ordered result.

Word encoding Handling computerized form of written language rests on processing of discrete symbols. One useful numerical representation of written text can be obtained by taking into account the sentential context in which the words occur. Before utilization of the context information, however, the numerical value of the code should not imply any order to the words. Therefore, it will be necessary to use uncorrelated vectors for encoding. The simplest method to introduce uncorrelated codes is to assign a unit vector for each word. When all different word forms in the input material are listed, a code vector can be defined to have as many components as there are word forms in the list. As an example related to Table 1 shown earlier, the color symbols of Table 2 are here replaced by binary numbers that encode them. One vector element (column in the table) corresponds to one unique color symbol.

0.250	0.235	0.215	1	0	0	0	0	...
0.165	0.042	0.042	0	1	0	0	0	...
0.222	0.184	0.135	0	0	1	0	0	...
0.210	0.105	0.030	0	0	0	1	0	...
0.255	0.127	0.080	0	0	0	0	1	...
0.184	0.134	0.011	0	0	0	0	0	1 ...
...	...	...	.	.	.	.	.	...

Table 2. A simple example of input data for the SOM algorithm in order to obtain map of symbols. The three first columns correspond to red-green-blue values and the rest of the columns are used to code the color symbols as binary values.

The component of the vector where the index corresponds to the order of the word in the list is set to the value "1", whereas the rest of the components are "0". This method, however, is only practicable in small experiments. With a vocabulary picked from a even reasonably large corpus the dimensionality of the vectors would become intolerably high. If the vocabulary is large, the word forms can be encoded by quasi-orthogonal random vectors of a much smaller dimensionality [47]. Such random vectors can still be considered to be sufficiently dissimilar mutually and not to convey any information about the meaning of the words. Mathematical analysis of the dimensionality reduction and the random encoding of the word vectors is presented in [47] and [21]. The random encoding can also be motivated from the linguistic point of view. The appearance of a word does not usually correlate with its meaning. However, it may be interesting to consider an encoding scheme in which the form of the words is taken into account to some extent. Motivation for such an experiment may stem, for instance, from the attempt to model aphasic phenomena.

Map creation and implicit categories The basic steps for creating maps of words are given in the following.

1. A unique random vector is created for each word form in the vocabulary.
2. All the instances of the word under consideration, so-called key words, are found in the text collection. The average over the contexts of each key word is calculated. The random codes formed in step 1 are used in the calculation. The context may consist of, e.g., the preceding and the succeeding word, or some other window over the context. As a result each key word is associated with a contextual fingerprint.
3. Each vector formed in step 2 is input to the SOM. The resulting map is labeled after the training process by inputing the input vectors once again and by naming the best-matching neurons according to the key word part of the input vector.

The averaging process is well motivated when the computational point of view is considered. The number of training samples is reduced considerably in the averaging.

The areas on a map of words can be considered as implicit categories or classes that have emerged during the learning process. Consider, for instance, Figure 2 in which some syntactic classes have emerged on a map of words. The overall organization of this map of words reflects syntactical categories. The context of the analysis has consisted of the immediate neighboring words. Single nodes can be considered to serve as adaptive prototypes. Each prototype is involved in the adaptation process in which the neighbors influence each other and the map is gradually finding a form in which it can best represent the input. Local organization of a map of words seems to follow semantic features. Often a node becomes labeled by several symbols that are synonyms, antonyms or otherwise belong to a closed class (see, e.g., [6, 19]). Often the class borders can be detected by analyzing the distances between the prototype vectors in the original input space [51, 52].

The prototype theory of concepts involves that concepts have a prototype structure and there is no delimiting set of necessary and sufficient conditions for determining category membership that can also be fuzzy. Instances of a concept can be ranked in terms of their typicality. Membership in a category is determined by the similarity of an object's attributes to the category's prototype. The development of prototype theory is based on the works by, e.g., Rosch [48] and Lakoff [29]. MacWhinney [32] discusses the merits and problems of the

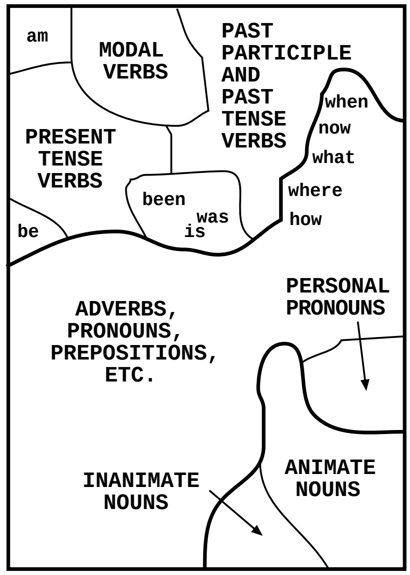

Fig. 2. Emergent implicit classes on a map of words. The input consisted of the English translations of fairy tales collected by the Grimm brothers. The 150 most frequent words were mapped. The area in the middle consists of words in classes of adverbs, pronouns, prepositions only in a partial order. Some of the individual words have been shown. Detailed results are presented in [13].

prototype theory. He mentions that prototype theory fails to place sufficient emphasis on the relations between concepts. MacWhinney also points out that prototype theory has not covered the issue of how concepts develop over time in language acquisition and language change, and, moreover, it does not provide a theory of representation. MacWhinney's competition model has been designed to overcome these deficits. Recently, MacWhinney has presented a model of emergence in language based on the SOM [33]. His work is closely related to the adaptive prototypes of the maps of words. In MacWhinney's experiments the SOM is used to encode auditory and semantic information about words. Also Gärdenfors' recent work (e.g., [9, 10]) is very closely related to the issue of adaptive prototypes. Mitra and Pal have studied the relationship between fuzziness, self-organization and inferencing [42] and the use of the SOM as a fuzzy classifier [41].

Handling Ambiguity The main disadvantage of the averaging process described earlier seems to be that information related to the varying use of single words is lost. However, it is entirely possible to use the SOM to cluster the contexts of a word to obtain information about the potential ambiguity of a word. Such a study has been conducted in [45]. Gallant [8] has presented a disambiguation method based on neural networks. In a study with similar objectives,

[50] used co-occurence information to create lexical spaces. The dimensionality reduction was based on singular value decomposition. An automatic method for word sense disambiguation was developed: a training set of contexts is clustered, each cluster is assigned a sense, and the sense of the closest cluster is assigned to the new occurrences. Schütze used two clustering methods to determine the sense clusters.

Related Work Miikkulainen has widely used the SOM to create a model of story comprehension. The SOM is used to make conceptual analysis of the words appearing in the phrases [37, 38, 40]. A model of aphasia based on the SOM is presented in [39]. The model consists of two main parts: the maps for the lexical symbols in the different input and output modalities, and the map for the lexical semantics. The SOM appears to be appealing model for aphasia. Consider, for instance, a situation in which the word "lion" is used instead of "tiger", i.e., a case of neighboring items in a semantic map. On the other hand, the use of "sing" instead of "sink" corresponds to a small error in a phonetic map.

Scholtes has used the SOM to parsing and several other natural language processing tasks such as filtering in information retrieval [49]. Several authors have presented methods for creating maps of documents rather than maps of words, e.g., [20, 30, 36]. The basic idea is to provide a visual display for exploration of text database. On the display two documents appear close to each other if they are similar in content. In [14, 15], a map of words is used as a filtering preprocessor of the documents to be mapped.

3.2 Modeling Gradience and Non-symbolic Representations

The world is continuous and changing, and, thus, the language is a medium of abstraction rather than a tool to create an exact "picture" of selected portions of the world. In the abstraction process, the relationship between a language and the world is one-to-many in the sense that a single word or expression in language is most often used to refer to a set or to a continuum of situations in the world. In order to be able to model the relationship between language and world, the mathematical apparatus of the predicate logic, for instance, does not seem to provide enough representational power. One way of enhancing the representation is to take into account the unclear boundaries between different concepts. Many names have been used to refer to this phenomenon such as gradience, fuzziness, impreciseness, vagueness, or fluidity of concepts.

The possibility of abandoning the predetermined discrete, symbolic features is worth consideration. However, a remark on the notion of 'symbol' may be necessary: the basic idea is to consider the possibility of grounding the symbols based on the unsupervised learning scheme. The symbols are used on the level of communication and may be used as the labels for the usually continuous multi-dimensional conceptual spaces. Symbols serve also as a means for compressing information.

Figure 3c outlines a scheme for “breaking up” the symbols. If suitable “raw data” are used, it may not be necessary to use any intermediate levels in facilitating interpretation. For example, de Sa [5] proposes a model of learning in a cross-modal environment based on the SOM. The SOM is used to associate input from different modalities (pictures, language). De Sa’s practical experiments use, however, input for which the interpretation of the features is given beforehand. Nenov and Dyer [43, 44] present an ambitious model and experiments on perceptually grounded language learning which is based on creating associations between linguistic expressions and visual images. Hyötyniemi [16] discusses the semantic considerations when dealing with mental models presenting three levels: features based on raw observations, patterns, and categories.

A mathematical framework for continuous formal systems is given in [31]; it can be based on, e.g., Gabor filters. Traditionally, the filters have to be designed manually. To facilitate automatic extraction of features, the ASSOM method [23, 26] could be used. For instance, the ASSOM is able to learn Gabor-like filters automatically from the input image data. It remains to be seen, though, what kinds of practical results can be acquired by aiming at still further autonomy in the processing, e.g., by combining uninterpreted speech and image input.

3.3 Contextuality, Subjectivity and Intersubjectivity

Considerable number of person years have been spent in coding the knowledge representations by means of traditional AI in terms of entities, rules, scripts, etc. It seems, however, that the qualitative problems are not satisfactorily solved by quantitative means. The world is changing all the time, and, perhaps still more importantly, the symbolic descriptions are not grounded. Symbol grounding, embodiment and their connectionist modeling is a central topic, e.g., in [53] and [46]. The contextuality of interpretation is easily neglected, being, nevertheless, a very commonplace phenomenon in natural language (see, e.g., [18]). The self-organizing map is a suitable method for contextual modeling: instead of handling symbols or variables separately the SOM can be used in associating them with the relevant context, or even in evaluating the relevancy of the context.

It seems that the SOM can be used in modeling the individual use of language: to create maps of subjective use of language based on examples, and, furthermore, to model intersubjectivity, i.e., to have a map that also models the contents of other maps (see, e.g., [12]). Two persons may have different conceptual or terminological “density” of the topic under consideration. A layman, for instance, is likely to describe a phenomenon in general terms whereas an expert uses more specific terms. However, in communication individuals tune in to each other’s language use. De Boer has simulated the emergence of realistic vowel systems in a population of agents that try to imitate each other as well as possible. The agents start with no knowledge of the sound system at all. Through communication a coherent vowel system emerges [2]. Similar thorough experiment using the SOM for conceptual emergence in an agent community still remains to be conducted.

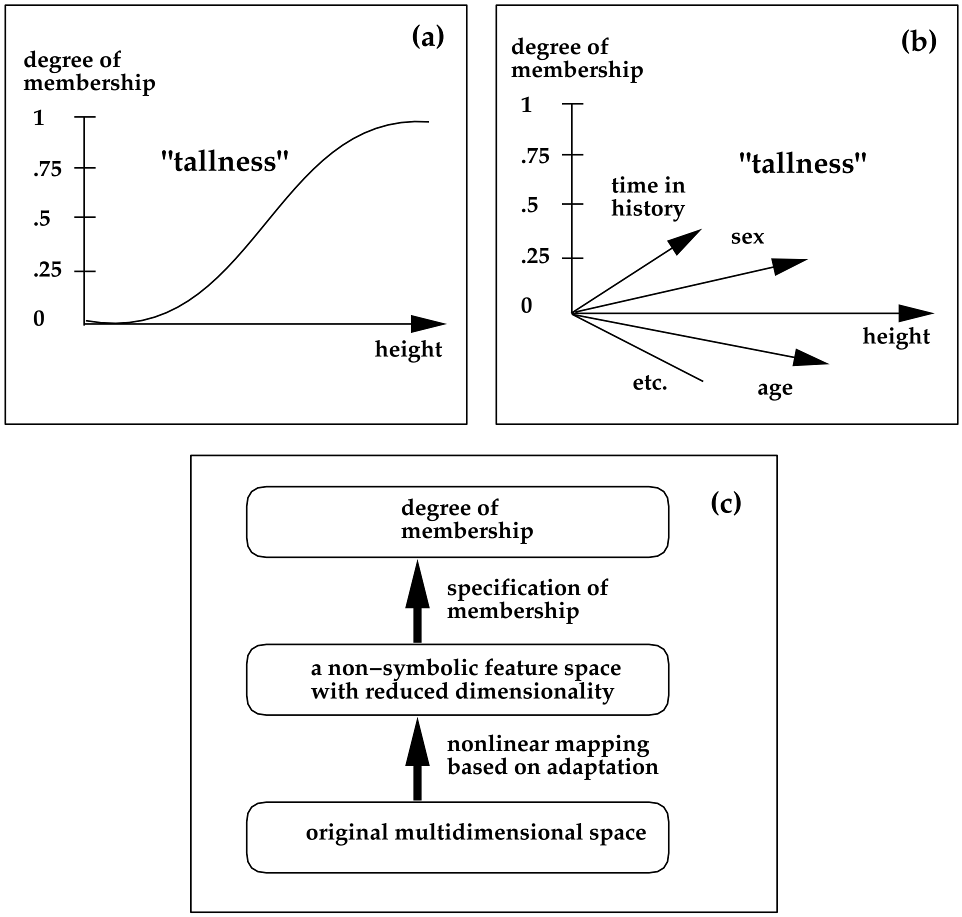

Fig. 3. Three stages of modeling continuity in the relation between linguistic expressions and the referenced continuous phenomena. In (a) a traditional view on fuzzy set theory is provided: the fuzziness of a single feature is represented by a curve in a one-dimensional space. The second alternative (b) points out the need to consider multidimensional cases: the degree of membership related to "tallness" of a person is not only based on the size of the person. The furthest scheme (c) is based on processing of continuous, uninterpreted "raw data" [15].

Honkela [12] proposes a model that adds a third vector element in addition to symbol part and context part: the specification of the identity of the utterer of the expressions. Thus, the network would be selective according to the "listener". This kind of enhanced map would provide a means for selecting the terms that can be used in communication. The detailed map could, for instance, include many expressions for different versions of a certain color among which a general term or a specific term would become selected based on the listener. This kind of use of the SOM provides a model of subjectivity and intersubjectivity: strictly speaking every "SOM-agent" has a symbol mapping of its own but the mappings are adapted through interactions to become similar enough to enable meaningful communication.

4 Conclusions

Natural language understanding as a field of artificial intelligence requires, for instance, means to model emergence of conceptual systems, intersubjectivity in communication, interpretation of expressions with contextual cues, and symbol grounding into perceptual spaces. If such means are available the methods can be considered to be relevant also in cognitive science and epistemology. In this paper, critical view on the traditional means to model natural language understanding and symbol processing has been given. Especially, the "purely symbolic methods", e.g., systems based on predicate logic or semantic networks, appear *not* to be sufficient to tackle the phenomena mentioned above.

As an alternative, the self-organizing map (SOM) has been considered in this paper. There are several results based on the SOM that cover relevant areas of the overall phenomena and thus the SOM seems to be a promising alternative for the more traditional approaches. There are, of course, related methods that can be used for similar purposes but the SOM covers several of the modeling requirements at the same time and, moreover, serves as a neurophysiologically grounded cognitive model. However, the basic SOM as such does not suit very well for processing structured information but one can combine the SOM with, e.g., recurrent networks to obtain better coverage of both structural and content-related linguistic phenomena (see, e.g., [34]). Moreover, the SOM principle can be generalized through adoption of principles of evolutionary computation which gives possibility of finding spatial order based, e.g., on functional similarity of potentially highly structured input samples [17, 28].

Acknowledgements

This work is mainly based on the results gained during my three-year stay at Academy Professor Teuvo Kohonen's Neural Networks Research Center in Helsinki University of Technology. I wish to thank Professor Kohonen for his most invaluable advice and guiding. I am also grateful to Dr. Samuel Kaski, Dr. Jari Kangas, Ms Krista Lagus, Dr. Dieter Merkl, Prof. Erkki Oja, Mr. Ville Pulkki, and many others.

References

1. Amari, S.-I.: A theory of adaptive pattern classifiers. *IEEC*, 16:299–307 (1967)
2. Boer, B. de: Investigating the Emergence of Speech Sounds. *Proceedings of IJ-CAI'99, International Joint Conference on Artificial Intelligence.* Dean, T. (ed.), Morgan Kaufmann, Vol. I, pp. 364–369 (1999)
3. Carpenter, G. and Grossberg, S.: A massively parallel architecture for a self-organizing neural pattern recognition machine. *Computer Vision, Graphics, and Image Processing*, 37:54–115 (1987)
4. Charniak, E.: *Statistical Language Learning.* MIT Press, Cambridge, MA (1993)
5. de Sa, V.: *Unsupervised Classification Learning from Cross-Modal Environmental Structure.* PhD thesis, University of Rochester, Department of Computer Science, Rochester, New York (1994)
6. Finch, S. and Chater, N.: Unsupervised methods for finding linguistic categories. I. Aleksander and J. Taylor, eds., *Artificial Neural Networks, vol. 2*, pp. II-1365-1368, North-Holland (1992)
7. Foerster, H. von: Notes on an epistemology for living things. *Observing Systems*, Intersystems publications, pp. 258-271 (1981)
8. Gallant, S. I.: A practical approach for representing context and for performing word sense disambiguation using neural networks. *ACM SIGIR Forum*, 3(3):293–309 (1991)
9. Gärdenfors, P.: Mental representation, conceptual spaces and metaphors. *Synthese*, 106:21–47 (1996)
10. Gärdenfors, P.: *Philosophy and Cognitive Science*, chapter Conceptual spaces as a framework for cognitive semantics, pp. 159–180. Kluwer, Dordrecht (1996)
11. Honkela, T. and Vepsäläinen, A.M.: Interpreting imprecise expressions: experiments with Kohonen's Self-Organizing Maps and associative memory. *Proceedings of ICANN-91, International Conference on Artificial Neural Networks*, T. Kohonen, K. Mäkisara, O. Simula and J. Kangas (eds.), North-Holland, vol. I, pp. 897-902 (1991)
12. Honkela, T.: Neural Nets that Discuss: A General Model of Communication Based on Self-Organizing Maps. *Proceedings of ICANN-93, International Conference on Artificial Neural Networks*, Amsterdam, Gielen, S. and Kappen, B. (eds.), Springer-Verlag, London, pp. 408-411 (1993)
13. Honkela, T., Pulkki, V. and Kohonen, T.: Contextual relations of words in Grimm tales analyzed by self-organizing map. *Proceedings of ICANN-95, International Conference on Artificial Neural Networks*, F. Fogelman-Soulié and P. Gallinari (eds), vol. 2, EC2 et Cie, Paris, pp. 3-7 (1995)
14. Honkela, T., Kaski, S., Lagus, T., and Kohonen, T.: *Newsgroup exploration with WEBSOM method and browsing interface.* Technical report A32, Helsinki University of Technology, Laboratory of Computer and Information Science, Espoo, Finland (1996)
15. Honkela, T.: *Self-Organizing Maps in Natural Language Processing.* PhD Thesis, Helsinki University of Technology, Espoo, Finland (1997) See *http://www.cis.hut.fi/~tho/thesis/*
16. Hyötyniemi, H.: On mental images and 'computational semantics'. *Proceedings of Finnish Artificial Intelligence Conference*, Finnish Artificial Intelligence Society, Espoo, pp. 199–208 (1998)
17. Nissinen, A.S., and Hyötyniemi, H.: Evolutionary Self-Organizing Map *Proceedings of EUFIT98: European Congress on Intelligent Techniques and Soft Computing*, pp. 1596-1600 (1998)

18. Hörmann, H.: *Meaning and Context.* Plenum Press, New York (1986)
19. Kaski, S., Honkela, T., Lagus, K., and Kohonen, T.: Creating an order in digital libraries with self-organizing maps. In *Proceedings of WCNN'96, World Congress on Neural Networks* (1996)
20. Kaski, S., Honkela, T., Lagus, K., and Kohonen, T.: WEBSOM—Self-Organizing Maps of Document Collections. *Neurocomputing*, 21:101-117 (1998)
21. Kaski, S.: Dimensionality reduction by random mapping: Fast similarity computation for clustering. *Proceedings of IJCNN'98, International Joint Conference on Neural Networks* (1998)
22. Kohonen, T.: Self-organizing formation of topologically correct feature maps. *Biological Cybernetics*, 43(1):59–69 (1982)
23. Kohonen, T.: The Adaptive-Subspace SOM (ASSOM) and its use for the implementation of invariant feature detection. Fogelman-Soulié, F. and Gallinari, P., editors, *Proceedings of ICANN'95, International Conference on Artificial Neural Networks*, volume I, pp. 3–10, Nanterre, France. EC2 (1995)
24. Kohonen, T.: *Self-Organizing Maps.* Springer, Berlin, Heidelberg (1995)
25. Kohonen, T., Hynninen, J., Kangas, J., and Laaksonen, J.: SOM_PAK: The Self-Organizing Map program package. Report A31, Helsinki University of Technology, Laboratory of Computer and Information Science (1996)
26. Kohonen, T., Kaski, S., Lappalainen, H., and Salojärvi, J.: The adaptive-subspace self-organizing map (ASSOM). *Proceedings of WSOM'97, Workshop on Self-Organizing Maps, Espoo, Finland, June 4-6*, pp. 191–196. Helsinki University of Technology, Neural Networks Research Centre, Espoo, Finland (1997)
27. Kohonen, T., and Hari, R.: Where the abstract feature maps of the brain might come fromo. Trends In Neurosciences, 22(3): 135-139 (1999)
28. Kohonen, T.: Fast Evolutionary Learning with Batch-Type Self-Organizing Maps. *Neural Processing Letters*, Kluwer, 9:153–162 (1999)
29. Lakoff, G.: *Women, Fire and Dangerous Things.* University of Chicago Press, Chicago (1987)
30. Lin, X., Soergel, D., and Marchionini, G.: A self-organizing semantic map for information retrieval. *Proceedings of 14th. Ann. International ACM/SIGIR Conference on Research & Development in Information Retrieval*, pp. 262–269 (1991)
31. MacLennan, B.: Continuous formal systems: A unifying model in language and cognition. *Proceedings of the IEEE Workshop on Architectures for Semiotic Modeling and Situation Analysis in Large Complex Systems, August 27-29, 1995, Monterey, CA* (1995)
32. MacWhinney, B.: *Linguistic categorization*, chapter Competition and Lexical Categorization. Benjamins, New York (1989)
33. MacWhinney, B.: *Cognitive approaches to language learning*, chapter Lexical Connectionism. MIT Press (1997)
34. Mayberry, M.R., and Miikkulainen, R.: SardSrn: A Neural Network Shift-Reduce Parser. *Proceedings of IJCAI'99, International Joint Conference on Artificial Intelligence.* Dean, T. (ed.), Morgan Kaufmann, Vol. II, pp. 820–825 (1999)
35. McCawley, J.D.: *Everything that Linguists have always Wanted to Know about Logic but were ashemed to ask.* Basil Blackwell, London (1981)
36. Merkl, D.: *Self-Organization of Software Libraries: An Artificial Neural Network Approach.* PhD thesis, Institut für Angewandte Informatik und Informationssysteme, Universität Wien (1994)
37. Miikkulainen, R.: *DISCERN: A Distributed Artificial Neural Network Model of Script Processing and Memory.* PhD thesis, Computer Science Department, University of California, Los Angeles, Tech. Rep UCLA-AI-90-05 (1990)

38. Miikkulainen, R.: *Subsymbolic Natural Language Processing: An Integrated Model of Scripts, Lexicon, and Memory.* MIT Press, Cambridge, MA (1993)
39. Miikkulainen, R.: Self-organizing feature map model of the lexicon. *Brain and Language*, 59:334–366 (1997)
40. Miikkulainen, R. and Dyer, M. G.: Natural language processing with modular neural networks and distributed lexicon. *Cognitive Science*, 15:343–399 (1991)
41. Mitra, S. and Pal S.: Self-organizing neural network as a fuzzy classifier. *IEEE Transactions on Systems, Man and Cybernetics*, 24(3):385–99 (1994)
42. Mitra, S. and Pal, S.: Fuzzy self-organization, inferencing, and rule generation. *IEEE Transactions on Systems, Man & Cybernetics, Part A [Systems & Humans]*, 26(5):608–20 (1996)
43. Nenov, V. I. and Dyer, M. G.: Perceptually grounded language learning: Part 1 - A neural network architecture for robust sequence association. *Connection Science*, 5(2):115–138 (1993)
44. Nenov, V. I. and Dyer, M. G.: Perceptually grounded language learning: Part 2 - DETE: a neural/procedural model. *Connection Science*, 6(1):3–41 (1994)
45. Pulkki, V.: Data averaging inside categories with the self-organizing map. Report A27, Helsinki University of Technology, Laboratory of Computer and Information Science, Espoo, Finland (1995)
46. Regier, T.: A model of the human capacity for categorizing spatial relations. *Cognitive Linguistics*, 6(1):63–88 (1995)
47. Ritter, H. and Kohonen, T.: Self-organizing semantic maps. *Biological Cybernetics*, 61(4):241–254 (1989)
48. Rosch, E.: *Studies in cross-cultural psychology, vol. 1*, chapter Human categorization, pp. 3–49. Academic Press, New York (1977)
49. Scholtes, J. C.: *Neural Networks in Natural Language Processing and Information Retrieval.* PhD thesis, Universiteit van Amsterdam, Amsterdam, Netherlands (1993)
50. Schütze, H.: Dimensions of meaning. *Proceedings of Supercomputing*, pp. 787–796 (1992)
51. Ultsch, A.: Self-organizing neural networks for visualization and classification. Opitz, O., Lausen, B., and Klar, R., editors, *Information and Classification*, pp. 307–313, London, UK. Springer (1993)
52. Ultsch, A. and Siemon, H.: Kohonen's self organizing feature maps for exploratory data analysis. *Proceedings of INNC'90, International Neural Network Conference*, pp. 305–308, Dordrecht, Netherlands. Kluwer (1990)
53. Varela, F. J., Thompson, E., and Rosch, E.: *The Embodied Mind: Cognitive Science and Human Experience.* MIT Press, Cambridge, Massachusetts (1993)
54. von der Malsburg, C.: Self-organization of orientation sensitive cells in the striate cortex. *Kybernetik*, 14:85–100 (1973)
55. Wermter, S., Riloff, E., and Scheler, G.: *Connectionist, Statistical and Symbolic Approaches to Learning for Natural Language Processing.* Springer Verlag, New York (1996)