# SELF-ORGANIZING MAPS IN NATURAL LANGUAGE PROCESSING

**Timo Honkela**

Helsinki University of Technology
Neural Networks Research Centre
P.O. Box 2200
FIN-02015 HUT, FINLAND

Thesis for the degree of Doctor of Philosophy to be presented with due permission for public examination and criticism in the Auditorium F1 of the Helsinki University of Technology on Friday, 12th of December 1997, at 12 o'clock noon.

ESPOO 1997

**Abstract**

Kohonen's Self-Organizing Map (SOM) is one of the most popular artificial neural network algorithms. Word category maps are SOMs that have been organized according to word similarities, measured by the similarity of the short contexts of the words. Conceptually interrelated words tend to fall into the same or neighboring map nodes. Nodes may thus be viewed as word categories. Although no a priori information about classes is given, during the self-organizing process a model of the word classes emerges.

The central topic of the thesis is the use of the SOM in natural language processing. The approach based on the word category maps is compared with the methods that are widely used in artificial intelligence research. Modeling gradience, conceptual change, and subjectivity of natural language interpretation are considered. The main application area is information retrieval and textual data mining for which a specific SOM-based method called the WEBSOM has been developed. The WEBSOM method organizes a document collection on a map display that provides an overview of the collection and facilitates interactive browsing.

# Contents

# PREFACE

I wish to express my deepest gratitude to Academy Professor Teuvo Kohonen without whom this work would not have been possible. As the inventor of the Self-Organizing Map he has created this area, and as the leader of the Neural Networks Research Centre he has gathered a stimulating and effective environment. I wish to thank professor Kohonen for his most invaluable advice and fatherly personal encouragement. His support has covered an overwhelming expertise in the technical and methodological issues as well as in the philosophical underpinnings of the field. Without his efforts from the 1960's till these days the concentrated and well-equipped research would not have been possible.

As a member of the WEBSOM team I have received great help and support. Central parts of this work are based on this collaboration. Therefore, I wish to show my deepest gratitude to Dr.Tech. Samuel Kaski, Prof. Teuvo Kohonen, and Ms Krista Lagus. I also wish to thank former colleagues Dr.Tech. Ari M. Vepsäläinen and Mr. Ville Pulkki for their efforts, and all the current and former colleagues in the Neural Networks Research Centre and in the Laboratory for Information Processing Science of Helsinki University of Technology, in VTT Information Technology, in VTKK Tietopalvelu, in Kielikone project of Sitra foundation, and in the Department of Information Processing Science of University of Oulu.

I wish to thank the prereviewers Prof. Fred Karlsson and Prof. Pasi Koikkalainen for their invaluable and constructive comments on the work and suggestions for improvements.

I have received a lot of valuable help, advice and comments during the years of research work from many current and former supervisors and colleagues, collaborators, and academic friends. I wish to express my warmest thanks to Ms Riitta Alkula, Dr. Harri Arnola, Mr. Jukka Honkela, Ms Sanna Hätönen, Dr. Heikki Hyötyniemi, Dr. Mauri Kaipainen, Dr. Jari Kangas, Prof. Fred Karlsson, Ms Taru Kuhanen, Mr. Aarno Lehtola, Prof. Seppo Linnainmaa, Dr. Kirsti Lonka, Ms Sini Maury, Prof. Erkki Oja, Ms Eeva Palosuo, Dr. Inga-Britt Persson, Mr. Jarno Raukko, Mr. Timo Sampolahti, and Prof. Olli Simula. I hope that the inevitable omissions will be forgiven.

Last but not least, I would like to thank Dr. Kaarlo Jaakkola and Dr. Jari Ahlberg as well as Shiatsu specialist Raija Koivisto for their efforts in ensuring my health recovery without which I would not have been able to write this thesis.

The thesis is dedicated to my parents: Maila, who left this world too early, Kaisa and Unto; and to Kristiina and Johanna whose patience I gratefully acknowledge; and especially to Catherine who has given me her understanding support and the deepest motivation for the work in its last phases.

I hope that this work provides insight into how it is possible to learn to understand, and, on the other hand, why it is not always easy to understand, although the same words are used. As a more straightforward consequence I also wish that the methodological results support development of applications that help people to communicate with each other efficiently by providing means for finding relevant information, and, in the long run, help in preventing unnecessary misunderstandings.

Helsinki, 2nd of December, 1997

*Timo Honkela*

## LIST OF PUBLICATIONS

The thesis consists of an introduction and the following publications:

1. Timo Honkela, and Ari M. Vepsäläinen (1991): 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.

2. Timo Honkela (1993): 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, Stan Gielen and Bert Kappen (eds.), Springer-Verlag, London, pp. 408-411.

3. Timo Honkela, Ville Pulkki, and Teuvo Kohonen (1995): 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.

4. Timo Honkela, Samuel Kaski, Krista Lagus, and Teuvo Kohonen (1996): Exploration of full-text databases with self-organizing maps. Proceedings of ICNN'96, IEEE International Conference on Artificial Neural Networks, Volume I, IEEE Service Center, Piscataway, NJ, pp. 56-61.

5. Samuel Kaski, Timo Honkela, Krista Lagus, and Teuvo Kohonen (1996): Creating an order in digital libraries with self-organizing maps. Proceedings of WCNN'96, World Congress on Neural Networks, September 15-18, San Diego, California, Lawrence Erlbaum and INNS Press, Mahwah, NJ, pp. 814-817.

6. Timo Honkela (1997): Comparisons of self-organized word category maps. Proceedings of WSOM'97, Workshop on Self-Organizing Maps, Espoo, Finland, June 4-6, 1997, pp. 298-303.

7. Timo Honkela (1997): Self-Organizing Maps of words for natural language processing applications. Proceedings of International ICSC Symposium on Soft Computing, Nîmes, France, September 17-19, 1997, ICSC Academic Press, Millet, Alberta, Canada, pp. 401-407.

8. Timo Honkela, Samuel Kaski, Teuvo Kohonen, and Krista Lagus (1997): Self-Organizing Maps of Very Large Document Collections: Justification for the WEBSOM method. Data Highways and Information Flooding, A Challenge for Classification and Data Analysis, I. Balderjahn and R. Mathar and M. Schader (eds.), Springer-Verlag, in press.

9. Timo Honkela (1997): Learning to understand – general aspects of using Self-Organizing Maps in natural language processing. Proceedings of the CASYS'97, Computing Anticipatory Systems, Liège, Belgium, August, 1997, in press.

# SUMMARY OF THE PUBLICATIONS AND THE AUTHOR'S CONTRIBUTION

In Publication 1, Kohonen's Self-Organizing Map (SOM) is used to model the impreciseness of natural language interpretation. The author designed the study and conducted the practical experiments related to the SOM. Ari M. Vepsäläinen provided his expertise of associative memories.

Publication 2 extends the principles presented in Publication 1 by introducing a model of a "communicating agent" that is based on the SOM. The basic idea is to provide a model of adaptation that leads to intersubjectivity in interpretation of natural language expressions and to coherence in communication.

Publication 3 deals with the analysis of contextual relations of words. The work was based on the principle of self-organizing semantic maps presented by Ritter and Kohonen in 1989 in *Biological Cybernetics*. In their article, artificially generated, syntactically correct and meaningful short phrases were used as the input to the SOM. For the publication, the method was refined and the input consisted of segments of text from a natural corpus. The author made the initial experiments, which then were continued by Ville Pulkki under the guidance of Prof. Teuvo Kohonen and the author. Ville Pulkki introduced the idea of using word-specific maps in the preprocessing, rather than calculating a single averaged context vector for each word in the vocabulary.

Studies on the WEBSOM full-text analysis method are described in Publications 4, 5, and 8. The basic system is presented in Publications 4 and 5. In the former, organization of the document collection was still supervised, i.e., the information of the group into which each document belonged was included in the input vector. This method was used in order to improve separation of the classes. Publication 5 presents a fully unsupervised method with improvements in the document encoding. The original idea of using a two-stage SOM architecture was due to the author. Samuel Kaski was the main originator of the idea of using table lookups for the efficient encoding of the documents and the subsequent convolution. The original document maps were limited to some 10000 documents at maximum. To facilitate considerably larger maps for up to one million documents, Professor Kohonen later suggested some new methods for shortcut computation. Many ideas and details, as well as implementations and experiments were developed jointly in a team consisting of Samuel Kaski, Teuvo Kohonen, Krista Lagus, and myself. It is not possible to give the full account of all the contributions of the individual team members. However, in general, the first of author's main contributions were related to the word category maps. Publication 8 then analyses the field of information retrieval and data mining of text collections. The author's idea was to use a two-level SOM, called WEBSOM, as a tool for intelligent information retrieval. This author's contribution ensued from his rather long experience in natural language processing and information retrieval that guided the selection and handling of the main themes.

Publication 6 is a rather specific study of the word category maps aiming at methodological fine-tuning. The effects of the various parameter selections were tested by the author using a map comparison method presented by Samuel Kaski and Krista Lagus at the ICANN'96 conference.

Publication 7 presents how the word category maps can be used in natural language processing (NLP) applications. The main emphasis is on the WEBSOM method, and the concrete results presented in the publication are based on the close collaboration of the WEBSOM team. Publication 7 also discusses the nature of the found clusters and the general principles of using the SOM, e.g., in graded classification, following the argumentations set by Ritter and Kohonen in their article in the journal *Biological Cybernetics* in 1989. The author has introduced the concepts of "emerging categories" and "adaptive prototypes" related to the SOM that were first presented in the International Cognitive Linguistics Conference in Amsterdam, July, 1997, the papers of which have not been published yet.

In Publication 9 the introduction outlines some limits of the traditional symbolic approaches such as relying on fixed categorizations, problems related to handling language change, and, moreover, subjectivity and context-sensitivity of interpretation. One chapter summarizes the basic principles of using the SOM in natural language interpretation. Finally, the last chapter contains some epistemological considerations.

In summary, the author's main contribution was the idea of the two-level SOM architecture where the first map clusters the words statistically according to their context, and the second map utilizes the frequency histograms clustered by the first map. The linguistic interpretations are also mainly the author's contribution.

# LIST OF SYMBOLS AND ABBREVIATIONS

| | |
|---|---|
| AI | Artificial intelligence |
| ASSOM | Adaptive subspace self-organizing map |
| HTML | Hypertext markup language |
| IR | Information retrieval |
| KDD | Knowledge discovery in databases |
| MT | Machine translation |
| NLP | Natural language processing |
| RGB | Red-green-blue |
| SOM | Self-organizing map |
| SQL | Structured query language |
| WWW | World wide web |

| | |
|---|---|
| $\mathbf{x}$, $\mathbf{x}_k$ | input vector (data item), $k$th input vector |
| $t$ | discrete-time index |
| $N$ | number of input vectors |
| $\mathbf{m}_i$ | $i$th model vector |
| $h_{ij}$ | neighborhood kernel in the SOM algorithm |

# 1 INTRODUCTION: AUTHOR'S MOTIVATION FOR THE WORK

The background of the present thesis is related to the personal experiences that the author gained in the *Kielikone* project funded by the Sitra foundation in the 1980's. The project aimed at development of a natural-language database interface (Jäppinen et al., 1988a; Jäppinen et al., 1988b). The idea was that a user could type in questions and commands in Finnish, and the system would transform them into formal database queries. The system consisted of multiple levels: morphological analysis of inflected Finnish word forms (Jäppinen et al., 1983; Jäppinen and Ylilammi, 1986)[1], disambiguation, syntactical analysis based on dependency grammar (Lehtola and Valkonen, 1986; Valkonen et al., 1987), semantical analysis that used a set of tree transformation rules to transform the dependency tree into a tree of predicate structures (Lehtola and Honkela, 1988; Honkela, 1989), and finally a two-stage database query formulation that first produced a system-independent query expression (Hyötyniemi and Lehtola, 1988), and finally then formed the query (Hyötyniemi and Lehtola, 1991). Special-purpose formalisms and metatools were developed during the project (e.g., Lehtola et al., 1988b; Lehtola et al., 1988a). The approach was mainly grounded on rule-based formalisms that proved to be practical in the tasks of syntactic analysis. A major problem arose in the semantical analysis. Three key findings have been relevant as the initiating problems that motivated the research reported in this thesis:

- When developing a system for disambiguation and seeing the results of the morphological analyzer, the importance of regarding ambiguity as a natural and frequent phenomenon became obvious. Some experience on machine learning was gained when a prototype system was developed that generalized disambiguation rules based on examples.

[1] Another practically complete model of Finnish morphology, the two-level model was developed by (Koskenniemi, 1983) in University of Helsinki. The two-level model is generally applicable over various languages and language families. Thus, prototypes of the two-level model have been implemented for over 30 languages. The most comprehensive implementations exist for Finnish, English, Swedish, Russian, Swahili, French, Arabic and Basque (Lindén, 1993). A language-independent formalism, Constraint Grammar (CG) has also been developed for syntactic analysis (Karlsson, 1990; Karlsson et al., 1995c; Karlsson et al., 1995a; Karlsson et al., 1995b). The recognition rate for a large English corpus, when parsing new unrestricted running text and after a morphological analysis by the two-level model, is approximately 98%, i.e., only 2 words out of 100 get the wrong syntactic code (Järvinen, 1994).

When the analysis concerning the structure and content of natural language expressions is considered, one key factor for quantitative problems is the combinatorial explosion caused by the contextuality of interpretation at the semantic level. Ambiguity further increases the necessity of taking the context into account in order to make rational interpretations[2]. In a 500 000 word sample, (Järvinen, 1994) reports an ambiguity rate of 16.4%, i.e., 16 words out of 100 have more than one morphological or syntactic interpretation.

- There exist several reasons for which a natural-language database interface is prone to fail. The system may fail because the user has misunderstood the scope of the system. For instance, relating to a company data base the user may ask what the most profitable companies as investment targets are. The failure may be caused by the misconception that the system is "intelligent" rather than considered just as a tool. No response may be available from the system because one of its components fails. Most often the failing component is the semantic analysis. One syntactic structure may correspond to dozens of different semantical schemes and to a vast number of detailed reference relations. For example, the genitive is coded in Finnish by an inflectional word form ending with 'n' and often including some changes in the word stem, whereas in English the genitive is of the form "of the house" or "house-'s". Developing a morphological analyzer for Finnish requires that one must model how the word stem is transformed in the different cases. In semantic analysis the quantitative problem is obvious: many different classes of interpretations for the single syntactic structure such as the genitive must be considered. The genitive may refer to owning ("Eva's car"), having a specific property ("John's age"), or an abstract relationship ("Catherine's husband"), etc. To be able to interpret what kind of relationship is connected to the genitive structure one must possess very detailed knowledge about every word being used in the actual expressions. As a simplification one may state that the genitive structure of "A's B" (or "B of A") can be detected by a single rule as a syntactic structure, whereas making an semantic analysis requires a vast amount of fine grained knowledge. Therefore, for instance, building a natural-language database interface has turned out to be perhaps an even more challenging task than developing a system for machine translation (MT) because in the latter, reasonable results are obtainable by morphological and syntactical analysis and generation. The so-called transfer phase

[2] Ambiguity may occur as homonymy, a word having two distinct meanings, as polysemy, a word having two related meanings, as vagueness, or structural ambiguity

is often used to transform the sentence structure descriptions into the corresponding structures (syntax trees or graphs) of the target language. Usually the results of the MT system need not be perfect to be useful. On the other hand, when a natural-language database interface is considered, the result of the transformation must be syntactically correct in order to obtain a correct formal expression in a database query language. In addition, it would be annoying if the interpretation of the original query or command were syntactically correct (in the sense of the query language) but semantically incorrect, because the user gets misleading results.

- When developing the semantic analysis component there arose also qualitative problems, such as the graded phenomena. For instance, the user may ask the system to show the *largest* companies in the database. The system has to decide whether the user would like to find the companies with the largest turnover, number of employees, some other indicators, or a combination of them. The subjectivity in interpretation became more apparent when the rules for semantic analysis were collected, and when the experiences of the test users were considered (Honkela, 1989).

Research work and experimentation of the use of morphological analysis and generation tools along with the full-text database system called *Minttu* provided experience on the themes of information retrieval (cf. Alkula and Honkela, 1992). The Minttu system was developed by the Finnish State Computer Center (VTKK, currently the Tieto Group) and it is widely used for storage and retrieval of administrative documents including the Finnish law. The need for using morphological analyzers and generators arises from the fact that Finnish is a highly synthetic language. Many words are formed from a word stem and inflectional or derivational suffices generated by morphological transformations. Therefore, for instance, finding occurrences of a word in texts is problematic. The main result was a system architecture for keyword-based information retrieval (IR) systems that use morphological analyzers and generators of Finnish word forms (Alkula and Honkela, 1992). The study also provided us insight into the general, well-known problems of IR. Searching for relevant documents from a very large collection has traditionally been based on keywords and their Boolean expressions. Often, however, the search results show high recall and low precision, or vice versa.

In summary, the main motivation for the research project presented in this thesis arose from the wish to develop a framework based on a novel paradigm and a practical application in which some of the difficulties mentioned above would be taken into account. The problems encountered in developing a natural-language database interface led into a re-evaluation of the status of the traditional artificial intelligence methodology and, specifically, that of natural language processing. Also many traditional philosophical underpinnings had to be questioned. Both quantitative and qualitative issues had to be considered. Firstly, the amount of knowledge needed in a successful large-scale NLP application is vast. Secondly, it seems that knowledge cannot be adequately modeled only by means of symbolic representation formalisms based on predicate logic. Further argumentation for these conclusions will be given in Chapter 3.

The experiences described above called for a completely new paradigm for approximate semantic analysis of natural language expressions. It turned out that the Self-Organizing Map formalism (Kohonen, 1995c) might provide means to cope with the qualitative and quantitative problems described earlier.

In this thesis, the Self-Organizing Map will be introduced in Chapter 2. Chapter 3 describes the principles and practices of using the SOM in natural language processing and interpretation[3], and Chapter 4 presents the use of the SOM in a particular application area, i.e., information retrieval and data mining of document collections.

[3]The term 'processing' is here used to refer to applications, whereas the term 'interpretation' emphasizes the cognitive point of view.

# 2 THE SELF-ORGANIZING MAP

The basic Self-Organizing Map (SOM) can be visualized as a sheet-like neural-network array (see Figure 1), the cells (or nodes) 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 the correct output (or actually the cell into which the input is mapped) for an input. In the basic version, only one map node (winner) at a time is activated corresponding to each input. The locations of the responses in the array 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 (Kohonen, 1995c).

The SOM was developed by Prof. Teuvo Kohonen in the early 1980s (Kohonen, 1981a; 1981b; 1981c; 1981d; 1982a; 1982b). The first application area of the SOM was speech recognition, or perhaps more accurately, speech-to-text transformation (Kohonen et al., 1984; Kohonen, 1988).

## 2.1 The Self-Organizing Map algorithm

Assume that some sample data sets (such as in Table 1[4]) have to be mapped onto the array depicted in Figure 1; 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 stochastic SOM algorithm performs a regression process. Thereby, 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 (Kohonen, 1995c). Moreover, a batch version of the SOM algorithm may also be used (Kohonen, 1995c).

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. 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

[4] Collected by Ben Chi (*http : //uacsc2.albany.edu/ ~bec/color.html*).

| | | | |
|---|---|---|---|
| 250 | 235 | 215 | antique white |
| 165 | 042 | 042 | 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 RGB (red-green-blue) values of the color shown in the rightmost column.

vector is most similar to the input vector in some metric (e.g. Euclidean) 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. The random initial state, two intermediate states, and the final map are shown in Figure 2.

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

$$\begin{aligned} \mathbf{m}_i(t+1) &= \mathbf{m}_i(t) + \alpha(t)[\mathbf{x}(t) - \mathbf{m}_i(t)] && \text{for each } i \in N_c(t), \\ \mathbf{m}_i(t+1) &= \mathbf{m}_i(t) && \text{otherwise,} \end{aligned} \tag{1}$$

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

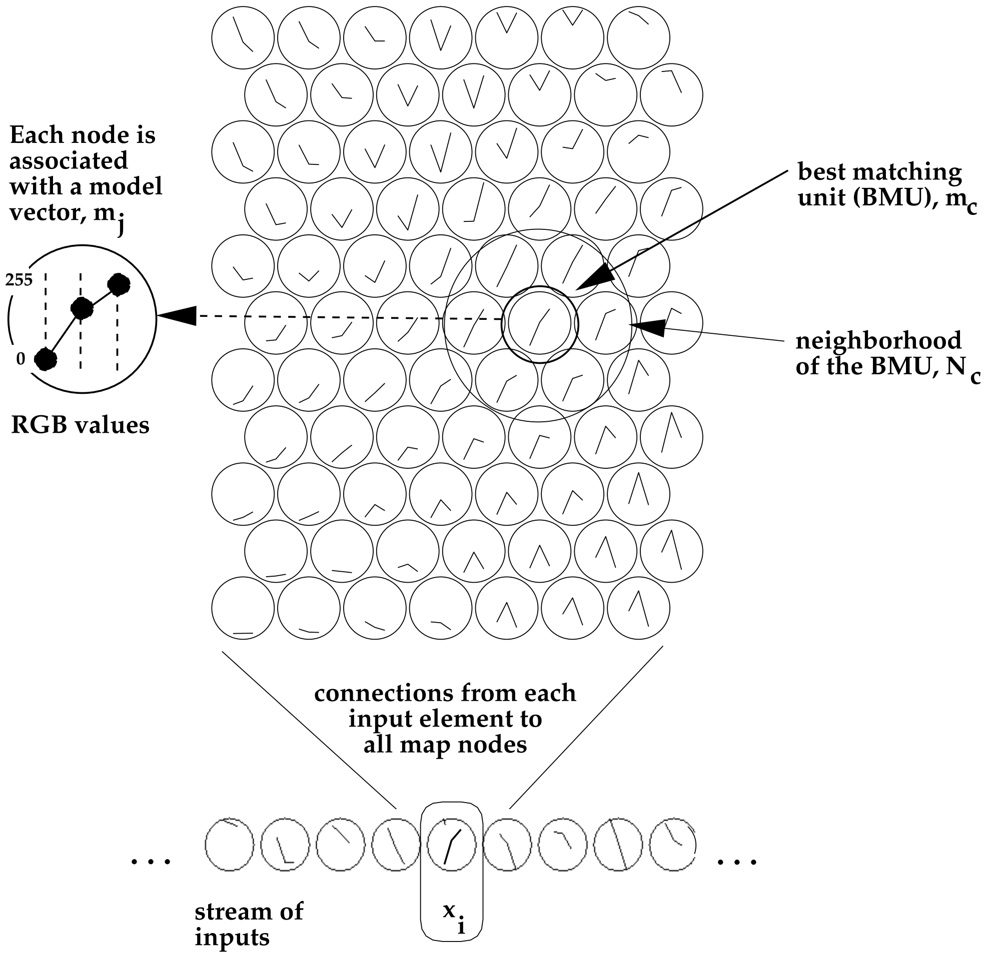


Figure 1: The basic architecture of the Self-Organizing Map (SOM). The input $x$ is fully connected to the array of map nodes which is most often and also in this illustration two-dimensional. Each map node, visualized as a circle on the grid, serves as a model, $m_i$, or to use another term, a prototype of a class of similar inputs. The line diagrams inside the circles denote the three RGB values of Table 1. For instance, the nodes on the lower left corner correspond to colors which have high values of all the components, i.e., red, green, and blue and therefore that corner contains very dark colors.

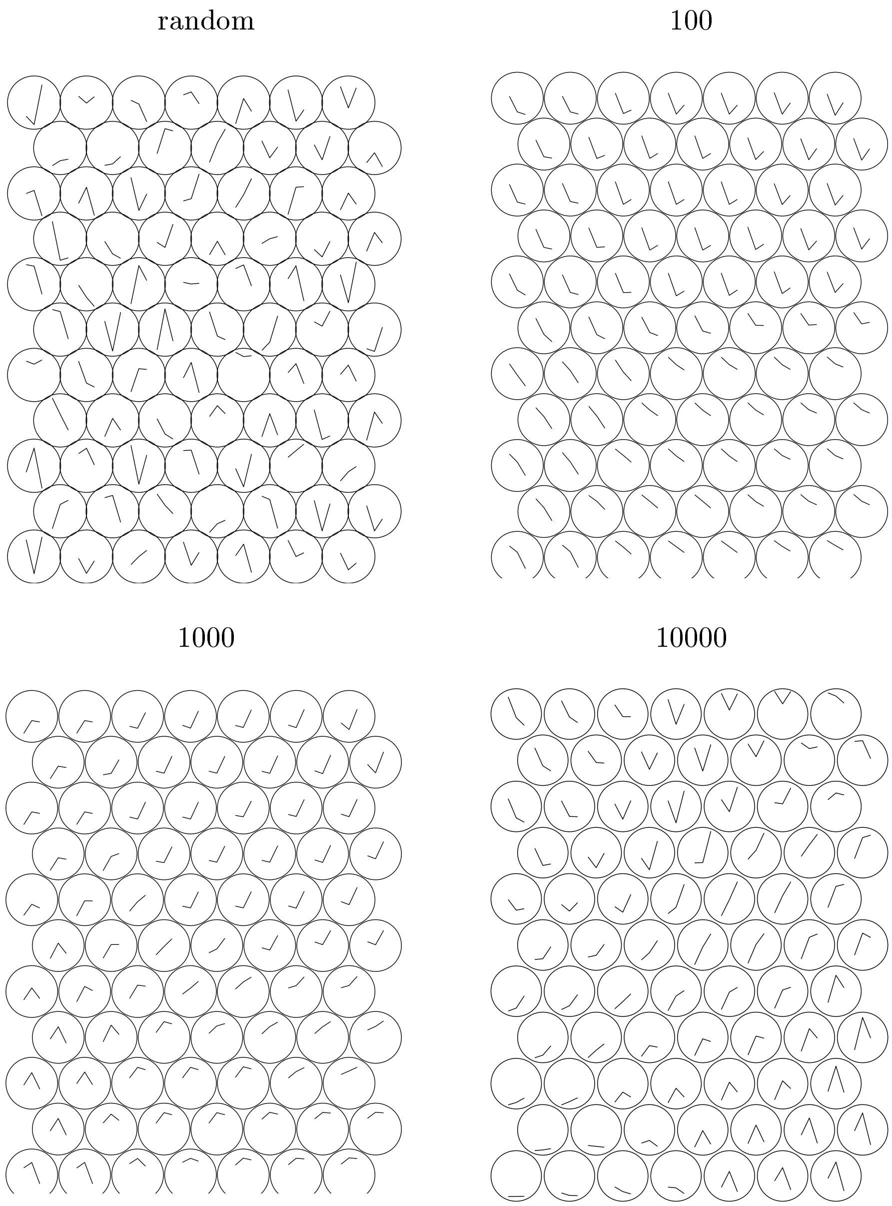


Figure 2: An example of the first steps of the ordering process of a $7 \times 11$ map. Each circle corresponds to a map node. Inside the circle the model vector consisting of the three RGB values from Table 1 is visualized. The initial value of the learning step size, $alpha_0$, was 0.2 and the neighborhood width was initially 5. The value of both parameters decreased linearly during the learning process. The values above the nodes indicate the number of learning steps.

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.

One method of evaluating the quality of the resulting map is to calculate the average quantization error over the input samples, defined as $\mathrm{E}\{\|\mathbf{x} - \mathbf{m}_c(x)\|\}$ where $c$ indicates the best-matching unit for $x$. After training, for each input sample vector the best-matching unit in the map is searched for, and the average of the respective quantization errors is returned.

The mathematical analysis of the algorithm has turned out to be very difficult. The proof of the convergence of the SOM learning process in the one-dimensional case was first given in (Cottrell and Fort, 1987). Convergence properties are more generally studied, e.g., in (Erwin et al., 1991; Erwin et al., 1992a; Erwin et al., 1992b; Horowitz and Alvarez, 1996; Flanagan, 1997). A number of details about the selection of the parameters, variants of the map, and many other aspects have been covered in the monograph (Kohonen, 1995c). The aim of this work is not to study or expound the mathematical and statistical properties of the SOM. The main point of view is to regard the SOM as a model of natural language interpretation, and explicate its use in natural language processing (NLP) applications, especially in information retrieval and data mining of large text collections, as introduced in the following chapter.

Most computations reported in the publications have been conducted using the SOM_PAK software package (Kohonen et al., 1996a).

## 2.2 Multiple views of the SOM

In the following, four particular views of the SOM are selected to serve as a basis for further inspection of the following chapters: 1. The SOM is a model of specific aspects of biological neural nets, 2. The SOM constitutes a representative of a new paradigm in artificial intelligence and cognitive modeling, 3. The SOM is a tool for statistical analysis and visualization, 4. The SOM is a tool for the development of complex applications.

1. 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. Some earlier artificial-neural-network models of self-organization have also been presented, e.g., by Shun-ichi Amari (Amari, 1967), Christoph von der Malsburg (von der

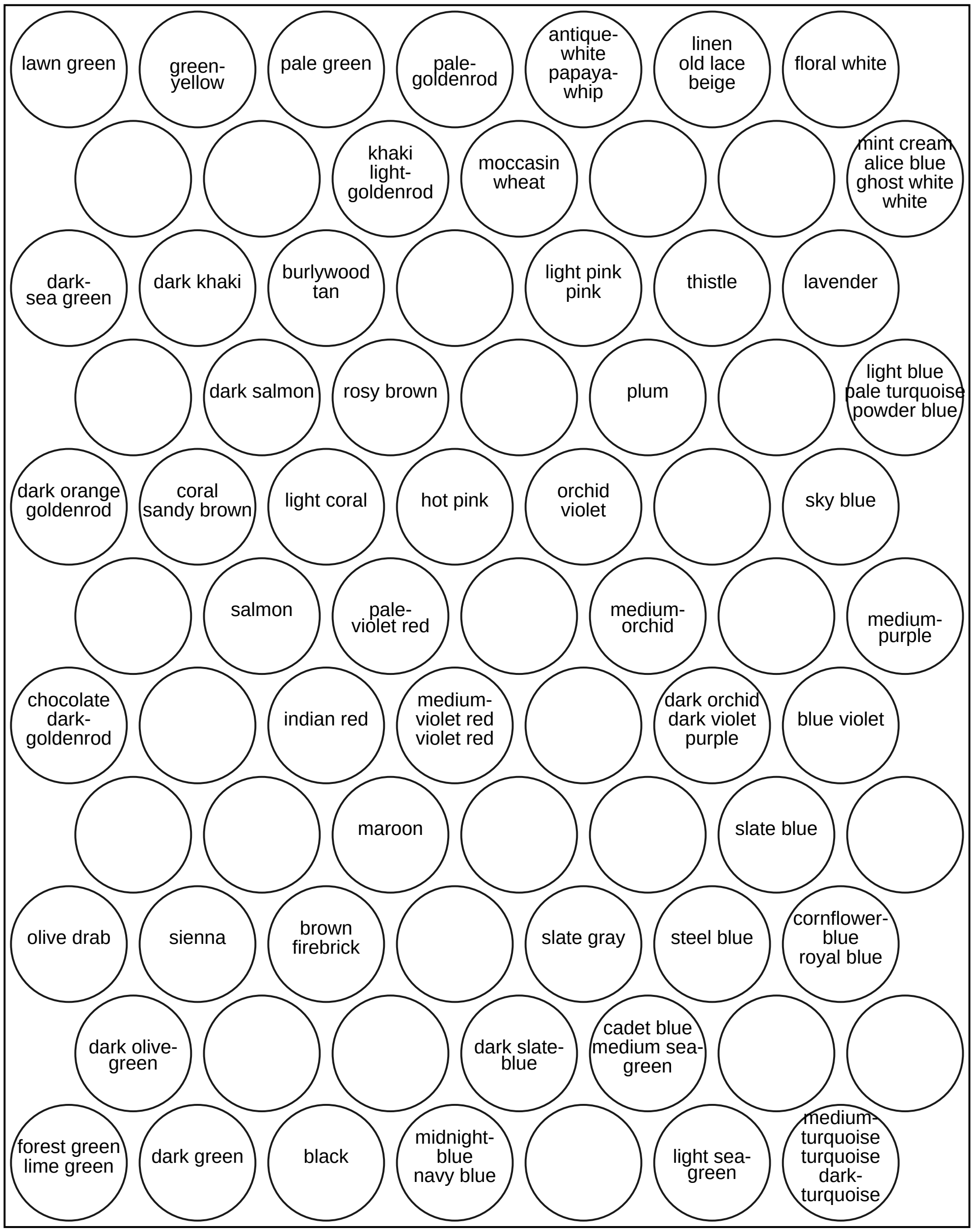


Figure 3: A map of colors based on their RGB values. The color symbols in the rightmost column of Table 1 as used in labeling the map. The best matching unit is searched for each input sample and that node is labeled accordingly.

Malsburg, 1973), and Gail Carpenter and Stephen Grossberg (Carpenter and Grossberg, 1987); however the SOM principle has turned out to produce the brain-like maps most efficiently.

Some experiments related to the localization of the linguistic functions are discussed in Chapter 3. The relationship between the SOM and its counterparts in neurodynamics is described in detail, e.g., in (Kohonen, 1992; 1993; 1996b). This issue, however, is only of indirectly related to this work, while the second and the fourth of these four threads are the main points of view to be considered.

2. The SOM can be viewed as a model of unsupervised (machine) learning, and as an adaptive knowledge representation scheme. In Publication 7 the relationship between the SOM (especially the word category map) and the semantic networks is considered. 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. Related to knowledge representation and learning, the cognitive and philosophical aspects are highly relevant. These issues will be considered in more detail in Chapter 3.

3. The SOM is nowadays often used as a statistical tool for multivariate analysis. The SOM is both a projection method which maps high-dimensional data space into low-dimensional space, and a clustering method so that similar data samples tend to be mapped to nearby neurons. From the methodological and computational point of view the mathematical and statistical properties of the algorithm can be considered (for instance, the time and space complexity, the convergence properties), as well as the nature of the input data (signals, continuous statistical indicators, symbol strings) and their preprocessing. There exist a number of variants of the SOM in which, e.g., the adaptation rules are different, various distance measures can be used, and the structure of the map interconnections is variable (Kohonen, 1995c).

4. The SOM is widely used as a data mining and visualization method for complex data sets. Application areas include, for instance, image processing and speech recognition, process control, economical analysis, and

diagnostics in industry and in medicine. A summary of the engineering applications of the SOM is given in (Kohonen et al., 1996c).

Some applications require efficient construction of large maps. Searching the best-matching unit is usually the computationally heaviest operation in the SOM. Using a tree-structured SOM, it is possible to use hierarchical search for the best match (Koikkalainen and Oja, 1990; Koikkalainen, 1994). In this method, the idea is to construct a hierarchy of SOMs, teaching the SOM on each level before proceeding the next layer. Another speedup method for making the winner search faster, based on the idea of Koikkalainen, is presented in (Kohonen et al., 1996b).

Most SOM applications use numerical data. The purpose of the present thesis is to demonstrate that statistical features of natural text can also be regarded as numerical features that facilitate the application of the SOM in NLP tasks.

# 3 SELF-ORGANIZING MAPS IN NATURAL LANGUAGE PROCESSING

Much of the formal and computational study of written language has centered around the structural aspects, to the detriment of semantics and pragmatics[5]. It seems likely that the methods available have been more suitable for the study of morphology and syntax. It also seems that the artificial neural network models can offer a promising methodology for dealing with these areas.

The connectionist study of morphology has been more limited in volume. However, already Koskenniemi (1983) discussed the relation between finite state automata used in his two-level model and neural networks. A widely referenced model of learning past tenses of English verbs was presented by Rumelhart and McClelland (1986). The aim was to demonstrate that (a) the model is able to capture the typical three-stage pattern of verb acquisition, (b) differences on performance on different types of regular and irregular verbs, and (c) the model responds appropriately to verbs that it has never seen before as well as to those included in the training set. Recently, Tikkala et. al (1997) have presented connectionistic models for simulating both normal and disordered word production as well as child language acquisition for Finnish.

In spite of the possibility to model linguistic phenomena in new ways, much of the connectionistic linguistic study has also mainly dealt with syntax and parsing. The relationship to main-stream generative linguistics is often close. Many models rely on the framework of well-known symbolic grammars (e.g., Faisal and Kwasny, 1990; Nakamura et al., 1990; Schnelle and Wilkens, 1990).

Connectionist approaches have been criticized, for instance, by claiming that a proper linguistic method should be able to represent constituent structures and to model compositionality (Fodor and Pylyshyn, 1988). Multiple answers can be given to this line of criticism. At the level of semantics and pragmatics, the contextuality may be considered to be more important: the final interpretation of an expression is determined in the context in which it appears. In addition, specific models that capture compositionality have been developed, e.g., RAAM, Recurrent Auto-Associative Memory (Pollack, 1990). Moreover,

[5]A definition of semantics and pragmatics that would be widely accepted is somewhat difficult to give. Levinson (1983) gives multiple possible definitions. One of the definitions is as follows: "Pragmatics is the study of the ability of language users to pair sentences with the contexts in which they would appear." He further states that this definition fits well with the definition of semantics according to which semantic theories are concerned with the recursive assignment of truth conditions to well-formed expressions of language. One general aspect of defining semantics is that it specifies the relation between linguistic expressions and the referents of the expressions.

the SOM is also able to represent implicit hierarchical structures. This issue will be handled in more detail in Section 3.5.

An overview of connectionist, statistical and symbolic approaches in NLP and an introduction to several articles is given by Wermter et al. (1996).

It may be asked why one should be interested in, e.g., the cognitive and philosophical point of view related to the NLP. The phenomena, as such, are naturally a relevant area of study. From the application point of view, the underlying principles of an approach used in constructing an NLP application may severely limit the chances to gain practical success. That may be considered to have happened in the traditional AI research and development. It remains to be seen, of course, how far one can reach by adopting the paradigm of radically connectionist[6] unsupervised learning devised, e.g., by the Self-Organizing Maps.

In conclusion, the aim of this work is to study the natural language processing in a rather reductionistic way when considered from the traditional point of view in which explicit theories of the phenomena are built. The aim is to study semantics and pragmatics to see which phenomena might be successfully modeled using the artificial neural network approach based on the Self-Organizing Maps, and to consider what the characteristics of the practical applications based on this approach are (Chapter 4). The next chapter introduces a model of automatic statistical lexical analysis based on the SOM.

## 3.1 Word category maps

Contextual information has widely been used in statistical analysis of natural language corpora. Charniak (1993) 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.

[6]Radically connectionist means here an approach in which artificial neural networks are used so that they are adaptive, and the intermediate representations are numerical and their interpretation can only be based on the adaptation process.

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

Table 2: The three first columns correspond to the RGB values and the rest of the columns are used to code the color symbols as binary values.

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. In all of the publications of this thesis, the algorithm used has been the SOM. The SOM does both vector quantization and clustering at the same time. Moreover, it produces a topologically ordered result.

Handling computerized form of written language rests on processing of discrete symbols. How can a symbolic input such as a word be given to a numeric algorithm? Similarity in the appearance of the words does not usually correlate with the content they refer to. As a simple example one may consider the words "window", "glass", and "widow". The words "window" and "widow" are phonetically close to each other, whereas the semantic relatedness of the words "window" and "glass" is not reflected by any simple metric.

One useful numerical representation 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[7]. 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.

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

[7]In this work, words are handled as the original word forms appearing in the text

“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 (Ritter and Kohonen, 1989). Such random vectors can still be considered to be sufficiently dissimilar mutually and not to convey any information about the meaning of the words. The influence of the random vector dimension on the pairwise deviance from the orthogonality is illustrated in Figure 4. Mathematical analysis of the random encoding of the word vectors is presented in (Kaski, 1998).

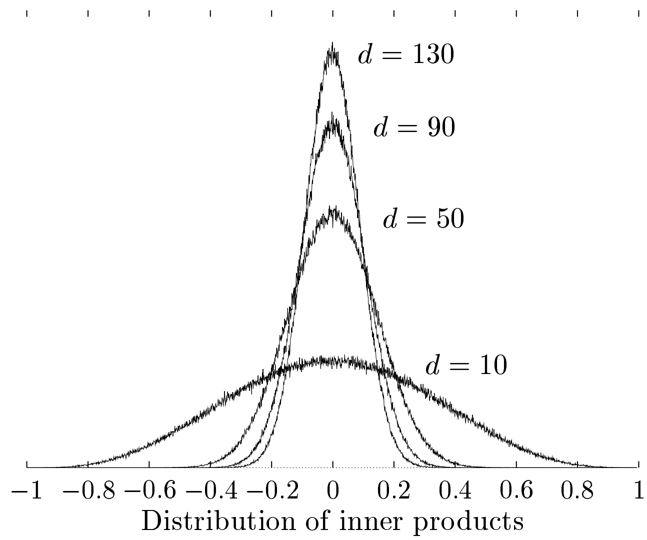


Figure 4: Distribution of the inner products between pairs of unit-norm random vectors for different dimensionalities ($d$). When the dimensionality increases the inner products become smaller and the vectors become more orthogonal and thus more suitable for encoding the words. The orthogonality will not be perfect but the generally small inner products contribute only small distortions in the similarity computations. The components of the vectors are normally distributed.

In the unsupervised formation of the word category map, each input $x$ consists of an encoded word form and its averaged context. Each word in the vocabulary is encoded as an n-dimensional random vector. In our experiments (see, e.g., Publication 3) $n$ has often been 90. A more straightforward way to encode each word would be to reserve one component for each distinct word but then, especially with a large vocabulary, the dimensionality of the vector would be computationally intractable. A mathematical analysis of the dimensionality reduction based on random encoding is presented by Ritter and Kohonen (1989).

The procedure of forming the word category map is presented in Publications 3, 4, 5, 7 and 9. In the following, the basic three steps are explained and motivated. Further methodological details are given in the publications.

1. A unique random vector is created for each word form in the vocabulary. The random encoding of the first step is motivated by computational and linguistic points of view. Linguistically, the appearance of the words does not usually correlate with their meaning. Computationally, if the words can somehow be encoded numerically, the ordered sets formed of them can also be compared mutually as well as with other numerical expressions.

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.

   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 main disadvantage seems to be that information related to the varying use of single words is lost.

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.

### Evaluating word category maps

Evaluating word category maps could, in principle, be based on calculating the quantization error of the map (see Chapter 2). However, in many applications the most important factor is the relative location of the words on the map. Therefore, measures for evaluation that take into account the word location are necessary. Kaski and Lagus (1996) present a method for comparing instances of SOMs. This method is used in Publication 6 to compare the results when the number of word forms on the map and the neighborhood type are considered.

The comparison method suits best to tasks in which the effect of parameter changes need to be evaluated.

Another possibility is to use a set of word lists as a basis of the evaluation. Each list consists of words that are expected to appear in nearby locations. Some preliminary tests have been made using this method. A practical measure that can be used in the word map evaluation is to make comparison in a further level of processing, e.g., when the document maps are created (see Chapter 4). The classification results on the document map can be used as the decision factor.

Often, it is practically impossible to create a map that satisfies all the expectations of a human viewer. The reason may be that either the input data simply does not contain necessarily detailed and balanced context information, or, the expectations may vary. The natural variation in multiple aspects of human linguistic behavior and how it can be modeled using the SOM is the main theme of the Section 3.4.

In Publications 7 and 9, various applications of word category maps are considered. The main application related to the current work is information retrieval and data mining of full-text document collections. The topic is introduced in Section 4.

In the following, a review of the research based on and related to the lexical SOMs is given. Thereafter, the potential of the word category maps in modeling three phenomena is considered: gradience or fuzziness in the relation between expressions and their referents (Section 3.3), subjectivity in generating and interpreting natural language expressions (Section 3.4), autonomous emergence of linguistically motivated implicit word classes or categories in a statistical analysis based on the contexts of the words (Section 3.5).

## 3.2 Other works based on and related to word category maps

As already mention earlier, the basic method for creating word category maps was introduced by Ritter and Kohonen (1989; 1990). Scholtes has applied the SOM in multiple NLP experiments. His main area is information retrieval (Scholtes, 1993).

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 (Miikkulainen, 1990a; 1990b; 1991; 1993a; 1993b; Miikkulainen and Dyer, 1991) . A model of aphasia based on the SOM is presented in (Miikkulainen, 1997). 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. A thorough framework for connectionist modeling of aphasia is presented in (Persson, 1995).

Scholtes has also used the SOM to parsing (Scholtes, 1992a; Scholtes and Bloemberger, 1992). An overall presentation is given in (Scholtes, 1993).

In (Schütze, 1992) cooccurence information is used to create lexical spaces. Here, the dimensionality reduction is based on singular value decomposition. In computing the context vectors, the window sizes are 1000 or 1200 characters rather than a fixed number of words. An automatic method for word sense disambiguation is 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 uses two clustering methods to determine the sense clusters.

Bookman's model for natural language interpretation, LeMICON (Bookman, 1994), combines symbolic and connectionist representations. LeMICON is a structured connectionist model that makes use of both connectionist and symbolic techniques to construct interpretations of text. The basic building blocks of LeMICON include a semantic memory and a working memory. Each concept in the semantic memory is represented by a vector of so-called "analog semantic features" of dimension 454. These 454 named features are based on a selection from the category structure of the Roget's thesaurus. Thus, the approach is more "classical" than the one adopted in this work. Bookman demonstrates how to perform automatic knowledge acquisition using a 7 million word database of Wall Street Journal articles. The semantic memory of the LeMICON system is constructed automatically based on an unsupervised learning algorithm.

## 3.3 Modeling gradience

When written language is analyzed, the static and discrete features of natural language are easily overemphasized[8]. For instance, the majority of knowledge representation formalisms used in the artificial intelligence and natural language processing research have been based on the tacit assumption of the

[8]Linell (1982) has written about the written language bias in the following way: "Our conception of linguistic behavior is biased by a tendency to treat processes, activities, and conditions of them in terms of object-like, static, autonomous and permanent structures, i.e., as if they shared such properties with written characters, words, texts, pictures and images. [...] In general, most of Western philosophy and science has been stuck with the metaphysical assumption that the world is made up of 'things' or 'objects'."

world consisting of entities and relations between these entities. The words in the language are considered to be "labels" of the entities[9]. The formalisms based on predicate logic contain predicates and their arguments (to model static structures of entities and their relations), logical connectives and quantifiers, and implications (to model rule-like phenomena and dependencies). Semantic networks and frame systems share the underlying ontological assumption. See, e.g., (Winston, 1984) for an overview of the "classical" AI techniques. Their influence on the models and metaphors used in cognitive science is also substantial, see, e.g., (Eysenck and Keane, 1990) for a textbook account.

Natural language is nowadays usually not viewed as a means for labeling the world but being an instrument by which the society and the individuals within it construct a model of the world. The world is continuous, and changing. 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. Thus, in order 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.

### Gradience and fuzzy logic

Perhaps the most widely used mathematical method of modeling gradience is fuzzy logic and fuzzy membership theory (Zadeh, 1965; Kosko, 1997). It is beyond the scope of this work to give a detailed description of fuzzy logic. For the purposes of the following discussion a simple example may be sufficient. Regarding the sentence "John is tall", the 'truth' of the sentence is a matter of degree. It is not possible to state a limit of tallness above which John would definitely be tall and under it not. In fuzzy logic the truth of a proposition is expressed by a value between 0 and 1 rather than being just 'true' or 'false' (see Figure 5a). Accordingly, an item may belong to a set in a fuzzy manner.

[9] In philosophy this view of the relationship between language and world was, among others, strongly proposed in the early works by Ludwig Wittgenstein. Perhaps his arrogance in Tractatus Logico-Philosophicus (saying that all the main problems are solved) was just premature.

The degree of membership of a set is also specified by a graded value from the real interval between 0 and 1. Often, however, a single dimension such as height is not enough: to state whether a person is tall one must also know her or his sex, age, genetic background, etc. (see Figure 5b). The degree of membership becomes, thus, a surface in a multi-dimensional space rather than a curve of one variable.

The status or necessity of fuzzy logic has been questioned, e.g., by many traditional logicians (see, e.g., Haack, 1996). Regardless of the philosophical arguments fuzzy logic has been found to be useful in constructing, for instance, practical control systems for cars, cameras, industrial processes, washing machines, etc. (Kosko, 1997).

Unsupervised learning can be used to learn the fuzzy descriptions of input data (Dickerson and Kosko, 1997). Using SOMs to learn descriptions of impreciseness of natural language is also the topic of Publication 1. The SOM is considered as a fuzzy classifier by Mitra and Pal (1994) (see also Mitra and Pal, 1996). Tirri (1991) and Leivian (1997) use the SOM to preprocess data before further processing by symbolic AI methods (rule-based expert systems, inductive learning).

### Beyond fuzzy logic: inherently non-symbolic representations

The possibility of abandoning the predetermined discrete, symbolic features is worth consideration[10]. The criticism towards fuzzy logic of going too far in rejecting the basic assumptions of traditional logics may even be reversed by claiming that fuzzy logic does not go far enough. Fuzzy logic is based on predetermined distinct features and therefore the descriptions created within it are not autonomous even if the membership clusters are found by using unsupervised learning.

Publication 9 discusses how the symbols could be "broken up". Figure 5c outlines a scheme for such an approach. If suitable "raw data" are used, it may not be necessary to use any intermediate levels in facilitating interpretation. For example, de Sa (1994) proposes a model of learning in a cross-modal environment based on the SOM. The SOM is used to associate input from different

[10] 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. What is the role of these symbols in further processing is left open in this work. However, the requirement of grounding the symbols is considered to be crucial.

modalities (pictures, language). De Sa's practical experiments use, however, input for which the interpretation of the features is given beforehand. Nenov and Dyer (1993, 1994) present an ambitious model and experiments on perceptually grounded language learning which is based on creating associations between linguistic expressions and visual images.

A mathematical framework for continuous formal systems is given by (MacLennan, 1995); 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 (Kohonen, 1995a; 1995b; 1996a; Kohonen et al., 1997a; 1997b) 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. The scope and time table of such study is, of course, far beyond the current work.

### Contextuality

The CYC project (Lenat et al., 1990) has collected a vast amount of knowledge to be used in large-scale natural language processing applications. Dozens 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 only adding more and more pieces of symbolic knowledge representations to a system. The world is changing all the time, and, perhaps still more importantly, the symbolic descriptions are not grounded[11]. The contextuality of interpretation is easily neglected, being, nevertheless, a very commonplace phenomenon in natural language (see, e.g., Hörmann, 1986). A model of contextuality in interpreting imprecise size expressions is presented in Publication 1.

The importance of context becomes evident when, e.g., ambiguity is considered: the interpretation of ambiguous expressions is based on their context. Pulkki (1995) presents means for modeling ambiguity using the SOM. The basic idea is to teach small "context maps" so that the main SOM can process the contextual information into clusters. Each model vector of the single-word maps corresponds to a particular "sense" of the word. The SOM is used to resolve ambiguity in (Scholtes, 1992b). Gallant (1991) has also developed a neural network based approach for word sense disambiguation. A model

[11] Symbol grounding, embodiment and their connectionist modeling is a central topic, e.g., in (Varela et al., 1993; Regier, 1995).

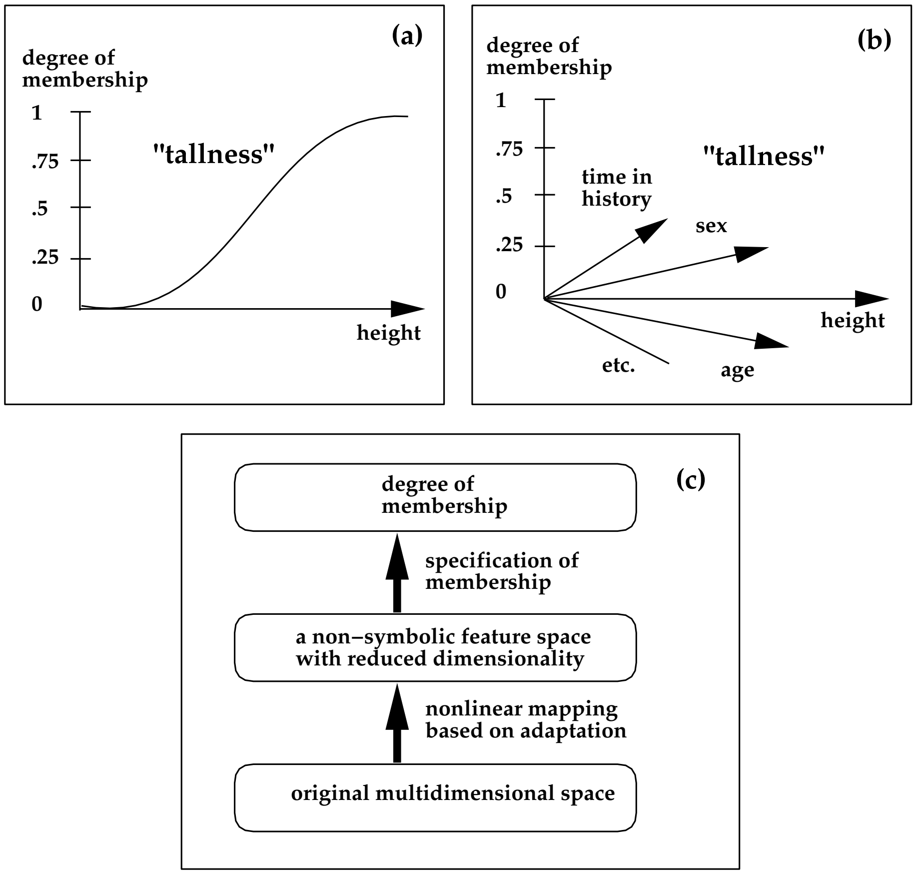


Figure 5: Three stages of modeling continuity in the relation between linguistic expressions and the referenced continuous phenomena. In (a) a traditional view on fuzzy set theory is provided: the fuzziness of a single feature is represented by a curve in a one-dimensional space. The second alternative (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".

for lexical disambiguation is presented by Mayberry and Miikkulainen (1994). The model is based on analyzing the frequencies of the contexts of ambiguous words. A recurrent network parser combines the context frequencies one word at a time, producing as the output the most likely interpretation of the current sentence.

A more traditional NLP approach is presented by Hirst (1987) with numerous illuminating examples of various types of ambiguity. Raukko (1994; 1997) discusses the linguistic phenomena of, or related to, ambiguity (polysemy, homonymy, vagueness) and proposes a view according to which the polysemous and vague words should not be considered as distinct entities but rather as patterns in continuous lexical space. This view is compatible with the approach adopted in this work.

## 3.4 Vocabulary problem and subjectivity

In information retrieval and collaborative systems the vocabulary problem is commonplace: people tend to use different terms to describe a similar concept. See (Chen, 1994) for a thorough analysis of the problem and for a proposal for solving it. In the following, the vocabulary problem is discussed shortly, and, next, the underlying issue of subjectivity in interpretation and generation of language is considered in more detail. Some examples are given regarding on how the SOM can be used in modeling the subjectivity of interpretation.

### Vocabulary problem

Furnas et al. (1987) have found that in spontaneous word choice for objects in five domains, two people favored the same term with less than 20% probability. Bates (1986) has shown that different indexers, well trained in an indexing scheme, might assign index terms for a given document differently. It has also been observed that an indexer might use different terms for the same document at different times.

Chen et al. (1995) reported that system-generated concept spaces outperformed human-generated thesauri in concept recall (i.e., the concept spaces provide more meaningful concept associations), but the human-generated thesauri are more precise in their suggested associations than those suggested by the system-generated thesauri.

**On subjectivity**

Moore and Carling (1988) state: "Languages are in some respect like maps. If each of us sees the world from our particular perspective, then an individual's language is, in a sense, like a map of their world. Trying to understand another person is like trying to read a map, their map, a map of the world from their perspective." 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. This is the main topic of Publication 2.

As an example related to the vocabulary problem, 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. This distinction is illuminated by the pair of color word maps in Figures 3 (in Chapter 2) and 6. The first map contains much more refined symbols for the colors.

Publication 2 proposes a model that adds a third vector element, i.e., the specification of the identity of the utterer of the expressions. Thus, the network would be selective according to the "listener". When, for instance, the maps in Figures 3 and 6 are considered, the enhanced map would provide a means for selecting which are the specific terms that can be used in communication with those having equal level of detailed knowledge and what is the scope of use of the general terms in the other cases. The detailed map contains, e.g., many expressions for different kinds of blue that would be uninterpretable when the second map is used. This naturally holds if the color is not present in the input. If the color is provided for both maps the less detailed model can be refined. Further discussion can be found in Publication 2.

When the input vector contains both context and symbol parts the degree of supervision can be moderated (controlled) using a single weight parameter (Ritter and Kohonen, 1990). In creating word category maps the symbol part of the input vector is normally scaled so that it does not significantly contribute to the ordering of the map. A corresponding case is presented in Figure 7. The different colors overlap, for instance, two instances of white ("white", "ghost white") can be found in the same node with an instance of very light blue ("alice blue"). In addition, the areas of orange, red, and brown are intertwined.

Figure 8 shows the result of an experiment in which the symbol part had 10 times higher weight than in the previous one. In both experiments, the colors

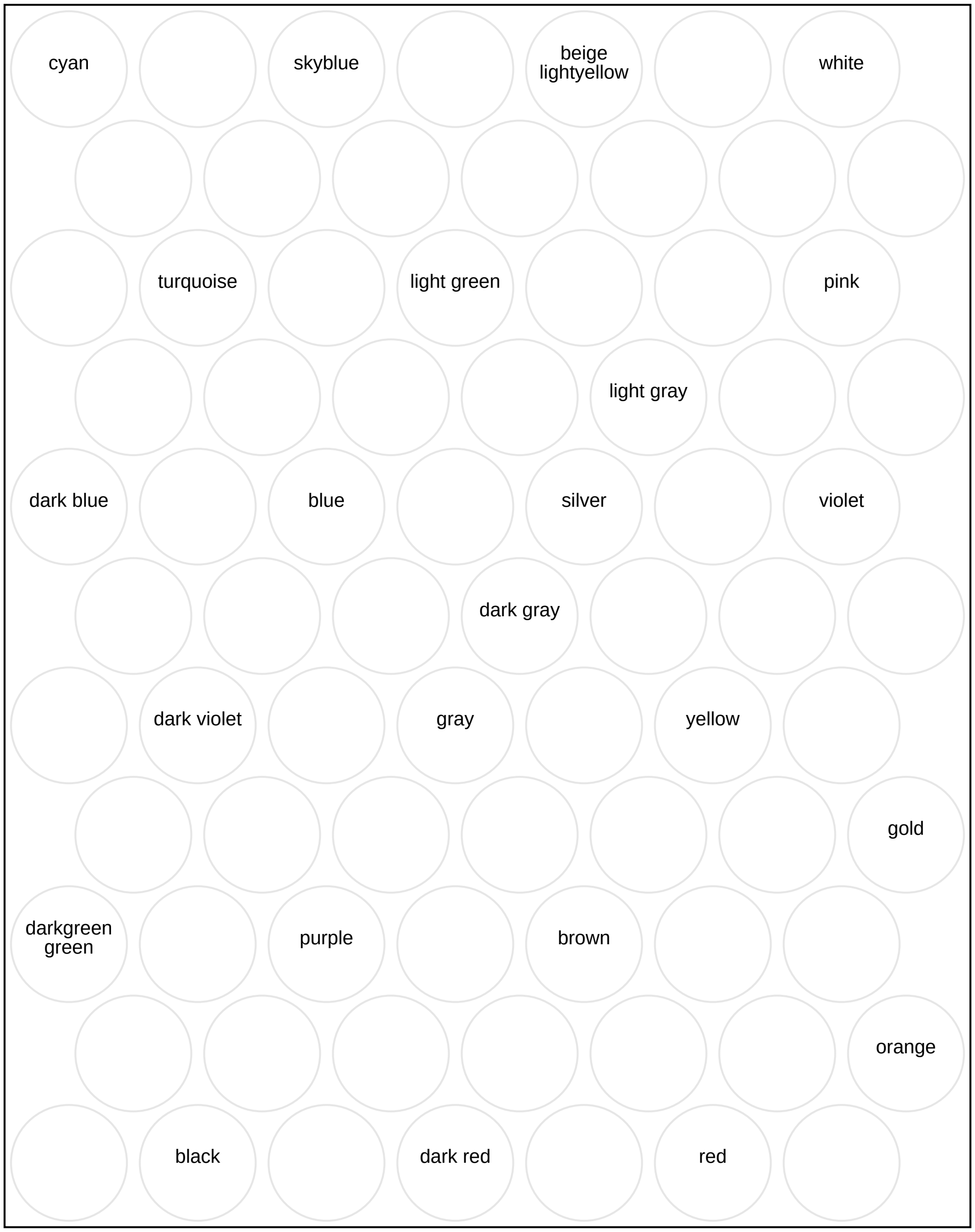


Figure 6: A map of a small number of colors based on their RGB values.

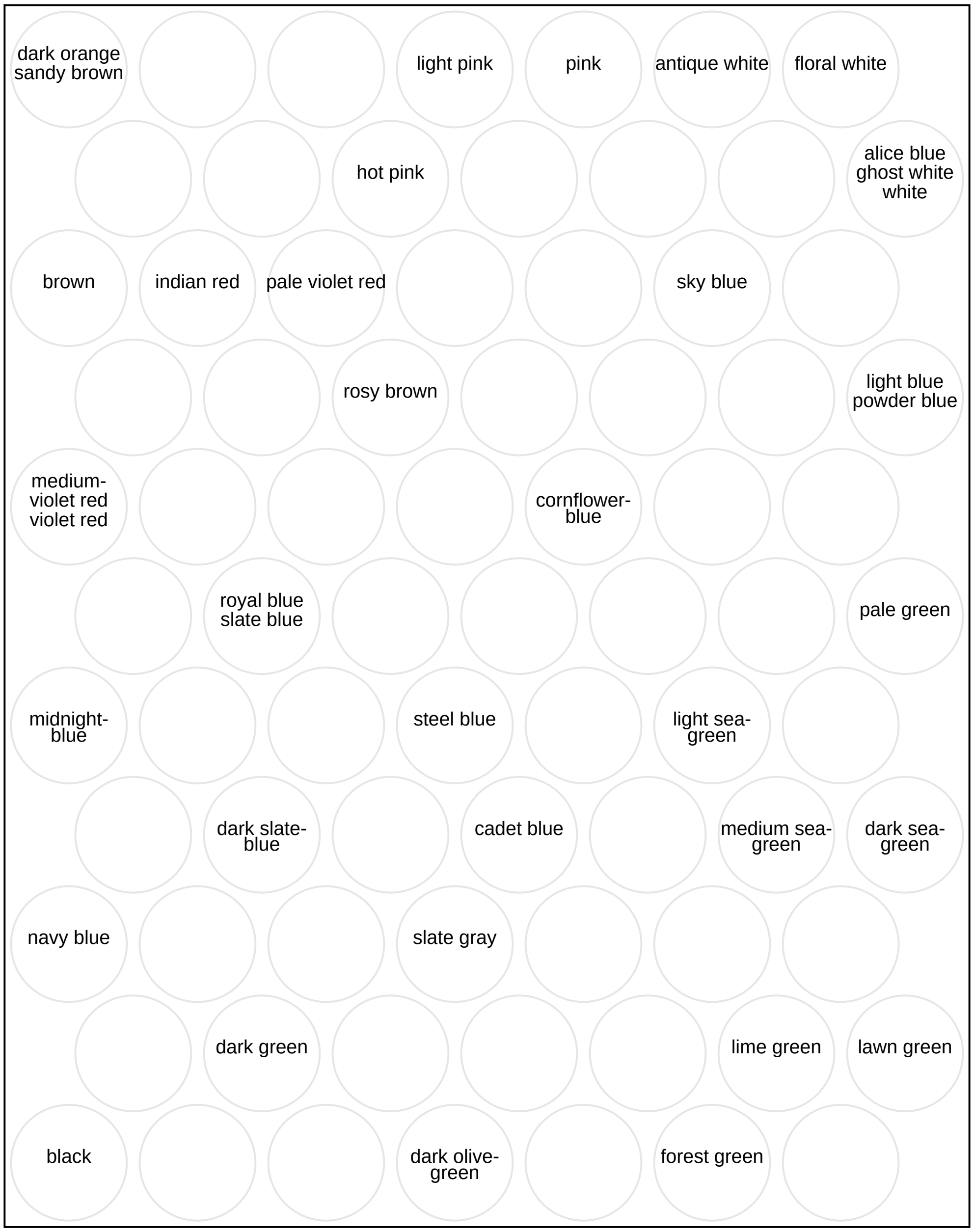


Figure 7: A map of colors. The input consists of RGB values (3 components) and of the symbol part that is encoded as a 9 dimensional binary vector scaled by the factor 0.01.

were classified into 9 classes based on the head word of the color. Thus, e.g., all the different instances of green shared the same symbol in the input vector. The difference of the maps reflects the classifying effect of the symbolic part[12].

As a summary, the potential of modeling subjectivity gives ground to applications in which the SOM is used to learn, e.g., the interests or the level of terminological knowledge of the user. Honkala (1997)[13] presents a model of browsing behavior by analyzing WWW user logs which provides an example of personalization (however, not based on linguistic analysis).

The possibility of modeling subjectivity can be contrasted with the traditional AI knowledge representation formalisms that provide very limited means for approaching this phenomenon. Brazdil and Muggleton (1991) use symbolic inductive inference as a means to learn to relate terms in a multiple agent communication as a kind of a model of intersubjectivity. They try to overcome language differences between agents automatically in a situation where the agents do not have the same predicate vocabulary. The system consists of a number of separate agents that can communicate. Each agent is able to perceive a portion of the given, possibly simulated world (in a symbolic form), accept facts and rules from another agent, formulate queries and supply them to another agent, respond to queries formulated by another agent, interpret answers provided by another agent, induce rules on the basis of facts, and integrate the knowledge. Their symbolic model of the environment is closely related to the model-theoretic approaches used in defining the semantics of formal languages. The inductive learning paradigm with symbolic representations does not, however, take into account that the meaning of an expression (queries, responses) in a natural domain is graded and changing, biased by the particular context, as discussed above.

### 3.5 Emergent categories and adaptive prototypes

The evidence for localization of linguistic functions in the brain is discussed, e.g., in (Caramazza et al., 1994). They cite studies in which category-specific deficits of selective impairment including proper names and geographical names have been reported (McKenna and Warrington, 1978). These findings and the maps presented in Publications 3 and 5 exhibit a striking resemblance. A number of studies is mentioned in (Caramazza et al., 1994) that show that

[12] A connection to the relation between thought and language may be considered: if the weight of the linguistic input is high enough, the overall organization is partly determined by it rather than being only based on the “perceptual” component.

[13] Not the present author.

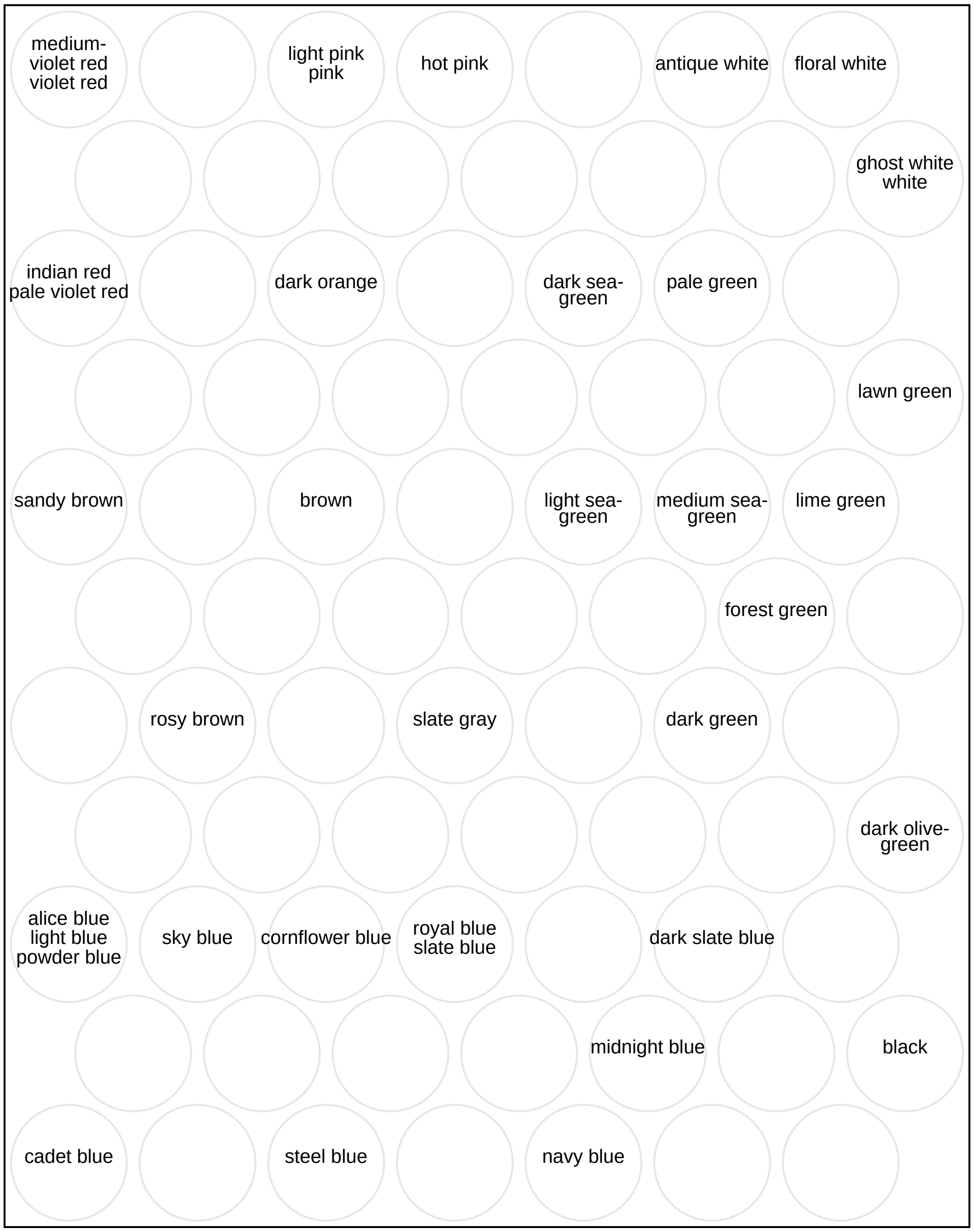


Figure 8: A map of colors. The input consists of RGB values and of the symbol part that is encoded as a 9 dimensional binary vector scaled by the factor 0.1. The structure of the map reflects the symbolic descriptions of the colors more than in the previous figure. Here, for instance, the area of the colors containing "blue" in their name forms a separate area of its own.

the category of living things can be selectively damaged while the non-living things have been spared, e.g., (Hillis and Caramazza, 1991). These findings match well with the structure of the map shown in Publication 3 in which the inanimate and animate nouns can be found in the areas of their own in the lower right corner of the map.

The areas on a word category map can be considered as implicit categories or classes that have emerged during the learning process. 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.

One classical approach for defining concepts is based on the idea that a concept can be characterized by a set of defining attributes. 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 (1977) and Lakoff (1987). MacWhinney (1989) discusses the merits and problems of the 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 (MacWhinney, 1997). His work is closely related to the adaptive prototypes of the word category maps, and to any maps of words that have been presented and discussed in Publications 1, 3, 4, 5, 6, 7, and 9. In MacWhinney's experiments the SOM is used to encode auditory and semantic information about words. Also Gärdenfors' recent work (e.g., Gärdenfors, 1996a; 1996b) is very closely related to the issue of adaptive prototypes.

# 4 APPLICATION OF SELF-ORGANIZING MAPS TO DATA MINING OF TEXTS

Data mining is a task that aims at identifying valid, novel, potentially useful, and understandable patterns in data (Fayyad et al., 1995). The general applicability of the SOM in data analysis and data mining becomes obvious from the proportion of the collection of 3014 references where the SOM has been used in tasks related to data mining[14]. The data mining use of the SOM has been extensively studied by Kaski (1997). In the following, some general aspects of data mining of texts are outlined.

In textual data mining and information retrieval the needs of the user may vary from looking for a specific piece of knowledge to a general wish to familiarize oneself with a subject area. The person's level of knowledge determines how well she or he is able select the terms and phrases in a specific search. In addition, the capability of using the methods and tools in a natural way influences the effectiveness of the search or exploration process. When evaluating information retrieval systems the user's needs are often considered rather narrowly, based on the relevance judgments of the results acquired from a query. The reason for this probably is that the relevance studies can be formalized easily. Overall user satisfaction is, of course, more important than any single relevance evaluation. To outline a broader view one may ask how well the user is aware of the domain of the inquiry, the document collection, and her or his information needs.

The document collections may vary in many respects, the most relevant of which for this thesis are: the size of the collection, the length of the documents, the style and type of content of the texts, and the possible additional information including classification of the documents and index terms.

The basic approaches for information retrieval and data mining for which a system may provide support are: (1) searching using keywords or key documents, (2) exploration of the document collection supported by organizing the documents in some manner, and (3) filtering. The keyword search systems can be automated rather easily whereas the document collections have traditionally been organized manually. The organization is usually based on some (hierarchical) classification schemes; each document is assigned manually to one or several classes. Another kind of organization can be provided by introducing associations between the documents manually. Filtering refers

[14] The bibliography of the Self-Organizing Map (SOM) and Learning Vector Quantization (LVQ) compiled in the Neural Networks Research Centre at Helsinki University of Technology is available at the WWW address *http://www.cis.hut.fi/nnrc/refs/*.

to discarding uninteresting documents from an incoming document flow.

## 4.1 Two traditional information retrieval methods

In the following, the traditional methods for information retrieval are presented to provide a background for the introduction of the SOM-based approach.

### Keyword search

One of the oldest methods of searching for texts that matches a query is to index all of the words (hereafter called the terms) that have appeared in the document collection. The query itself, typically a list of appropriate keywords, is converted into a list of terms. This list is compared with the term list of each document to find matching documents. Various Boolean logic expressions can be formulated to control the breadth of matching. For instance, instead of a simple list of terms, the query can be based on a Boolean expression of the terms. The following three basic problems make it difficult to apply Boolean logic to text retrieval (see, e.g., Salton, 1989). (1) Recall and precision[15] of retrieval is sensitive to small changes in formulation of a query. For Boolean queries there is no simple way of controlling the size of the output, and the output is not ranked in the order of relevancy. (2) Considering the results of a query it is not known what was *not* found, especially if the document collection is unfamiliar. (3) If the domain of the query is not known well it is difficult to formulate the query, i.e., to select the appropriate key words.

It is a very basic problem in the text document management that different words and phrases are used for expressing similar objects of interest. Natural languages are used for the communication between human beings, i.e., individuals with varying background, knowledge, and ways to express themselves. Therefore, if information retrieval is based only on the keywords as they appear in the text some interesting documents may not be discovered, or misleading documents may be returned. The traditional information retrieval methods have limited possibilities to tackle this phenomenon. Clearly, the information retrieval systems might therefore benefit from taking into account the contexts of the words and deriving relations of the words based on the contextual information.

One solution suggested for the selection of a vocabulary or of keywords is

[15] The two standard measures of retrieval effectiveness are precision, the number of relevant retrieved documents over the total number of retrieved documents, and recall, the ratio of relevant retrieved documents to the total number of (known) relevant documents.

the use of thesauri. The systems could return not only the documents that match the query but also documents that match some synonyms, found in the thesaurus. However, if a thesaurus is used, the system will often return a larger number of irrelevant documents. Another problem in using a thesaurus that it may be difficult to create and maintain manually.

### Vector space model

In the automatic recognition of items such as visual objects or segments of speech, one often represents the item as a real pattern vector of a number of feature values. One of the oldest pattern recognition methods is to compare the representations in some metric, e.g., the Euclidean metric, or based on similarity functions such as direction cosines between the pattern vectors.

The vector space model has been proposed by (Salton et al., 1975; Salton, 1989) to document collection analysis. It constitutes an alternative approach to the use of the Boolean algebra to combine the terms in the queries. In the vector space model, both the stored documents and the user's queries are represented by vectors where each dimension corresponds to a term. The value of a component, called a weight, may be a function of the frequency of the occurrence of the term in the document, and of how important the term is considered to be. Documents and queries are then compared based on some similarity functions between the vectors. Given a query, each document can be assigned a similarity score which indicates the quality of the match between the query and the document. The scores can be ranked and displayed to the user to provide the documents in the intended order of relevancy. Also word counts, i.e. term frequencies, can be used to measure differences of emphasis between documents. A major problem in the vector representation of documents is that in the representation all pairs are considered equally similar. Semantical relationships between the terms are not taken into account.

## 4.2 Requirements for information retrieval and textual data mining systems

A successful system should be able to tackle the most important problem areas of information retrieval and textual data mining that were discussed in Section 4.1. In addition, it should be possible to use the system in different ways to meet as many of the different needs as possible. In the following a short summary of the basic requirements is given. The system should:

- support all of the basic functions: searching, exploration, and filtering,
- require as little human intervention as possible in order to enable processing of very large document collections,
- in searching, provide the results in the order of relevancy and offer means for exploiting what was left out from the result,
- support exploration by giving views on the document collection,
- degrade gracefully even if the documents are poorly written,
- be computationally feasible, and
- take into account, in a general way, the inherent features of natural language, and adapt to the specific use of the system.

A thorough overview of connectionist models in information retrieval is given in (Doszkocs et al., 1990). The connectionist models, often combined with the principles of the vector space model, appear to be a promising alternative for the traditional keyword-based methods. In the following, a novel SOM-based method for information retrieval and textual data mining, called WEBSOM, is introduced. The WEBSOM aims at satisfying the requirements listed above.

## 4.3 WEBSOM - document maps based on word category maps

### SOM-based approaches

To produce maps that display similarity relations between document contents a suitable statistical method must be devised for encoding the documents. Perhaps the most straightforward method is to encode the documents based on the words or words forms they contain. In an early study, a small map of scientific documents was formed based on the word forms in their titles (Lin et al., 1991; Lin, 1992). Later also Lin has extended his method to full-text documents (Lin, 1997). When using the vector space model, the general procedure for converting documents to vectors can be summarized as three steps (modified from Lin, 1997):

1. Identify a vocabulary from the document collection (based on the word forms in the document titles, in the abstracts, or in the full texts).
2. Delete some common words using a stop list. Remove also the most and least frequently occurring terms.

3. Index the document collection on the basis of the remaining list. The components of a document vector can consist of:
   (a) binary digits–the value depends on whether the corresponding term appears in the document or not;
   (b) weights based on the frequency of occurrence of the term in the document;
   (c) weights based on the term frequency and the inverse document frequency. The inverse document frequency is the inverse of the number of documents in which the term occurs (Salton et al., 1975; Salton, 1989).

Scholtes has developed, based on the SOM, a neural filter and a neural interest map for information retrieval (Scholtes, 1993). In addition to encoding the documents based on their words, Scholtes has used character n-grams[16] in the encoding. This approach has also been adopted in (Hyötyniemi, 1996). Merkl (1993; 1994; 1995a; 1995b; 1997; Merkl et al., 1994) has used the SOM to cluster textual descriptions of software library components based on similar principles. Some of the experiments have been based on the use of hierarchical SOMs (see Chapter 2). Document maps have also been created by Rozmus (1995) and Zavrel (1995; 1996). The AI Group lead by Hsinchun Chen in the MIS Department at the University of Arizona report that they have used the SOM in their "ET-Map" in categorizing the content of Internet documents to aid in searching and exploration[17].

### The WEBSOM method

By virtue of the Self-Organizing Map algorithm, documents can be mapped onto a two-dimensional grid so that related documents appear close to each other. If the word forms are first organized into categories on a *word category map*, an encoding of the documents can be achieved that explicitly expresses the similarity of the word meanings. The encoded documents may then be organized by the SOM to produce a *document map*. The visualized document map provides a general view of the document collection.

The basic processing architecture of the WEBSOM method is presented in Figure 9. The details of the method are described in Publications 4 and 5. An overview is given below.

[16] N-gram is a sequence of n characters.
[17] A demonstration of the ET-Map is available at the WWW address *http://ai.bpa.arizona.edu/ent/*

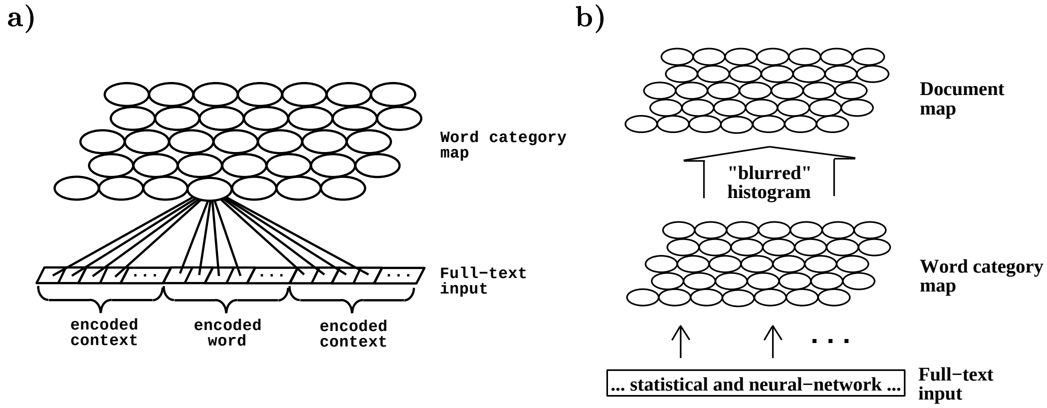


Figure 9: The basic two-level WEBSOM architecture. a) The word category map first learns to represent relations of words based on their averaged contexts. The word category map is used to form a word histogram of the text to be analyzed. b) The word category histogram is then used as input to the second SOM, the document map.

The documents are encoded by mapping their text, word by word, onto the word category map whereby a histogram of the "hits", a word category histogram, on it is formed. The histogram is a kind of "fingerprint" of the document. To reduce the sensitivity of the histogram to small variations in the document content the histograms can be smoothed using, e.g., a Gaussian convolution kernel (see Publication 4)[18]. This is possible since the word category map is spatially ordered according to semantic similarity relations between the words. The document map is then formed by the SOM algorithm using the word category histograms as inputs.

The WEBSOM has two possible modes of operation: unsupervised and supervised. In the former, clustering of the word category histograms of arbitrary text files is made by a normal Self-Organizing Map, whereby no class information about the documents is utilized. The organization of the documents is simply based on the analysis of the raw texts. In the supervised mode of operation, separation of the document groups of the document map is enhanced by utilizing auxiliary information about in which topic category the documents belong to. A closer study of using the supervised-SOM principle (Kohonen et al., 1984; Kohonen, 1995c) in the WEBSOM method and experimental results can be found in Publication 4. The first results based on

[18]The effect of this smoothing has, however, been found to be rather small at least in some applications (Kaski, 1997a)

unsupervised learning are presented in Publication 5 as well as in the technical report (Honkela et al., 1996).

### Browsing

A WWW-based browsing environment has been developed for supporting the exploration of document collections. The browsing tool provides a view of the map. First, the user may zoom at any map area of the main map by pointing to the map image to view the underlying document space in more detail. The clustering between the neighboring map nodes has been shown using different shades of gray (see, e.g., Ultsch and Siemon, 1990; Ultsch, 1993; Merkl and Rauber, 1997): if the model vectors are similar, light shades are used. Otherwise the dark shade indicates that there is a jump in neighboring model vectors. The WEBSOM browsing interface is implemented as a set of HTML[19] documents that can be viewed using a graphical WWW browser. The interface and its use in exploring document collections is considered in detail in (Lagus et al., 1996a; 1996b; 1996c). An example of a document map is shown in Figure 10.

A map of conference abstracts is presented in (Lagus, 1997). The use of the SOM in data mining of both numerical and textual data collections is considered in detail in (Kaski, 1997b). General aspects of using the document maps are presented in Publication 8.

### Recent developments of the WEBSOM

Since the first results the WEBSOM method has undergone significant developments. The developments, some of which are still unpublished are listed below.

- Methods that enable computation of very large maps have been developed (Kohonen et al., 1996b; Kohonen, 1997). The largest map so far contains over one million documents[20].

- The quality of the large document maps has been improved by enlarging the word category maps considerably. The overloading problem of word category maps, i.e., the problems caused by each map node containing too large a number of words is discussed in Publication 6.

[19] HyperText Markup Language

[20] The demonstration can be seen in the WWW address *http://websom.hut.fi/websom/*.

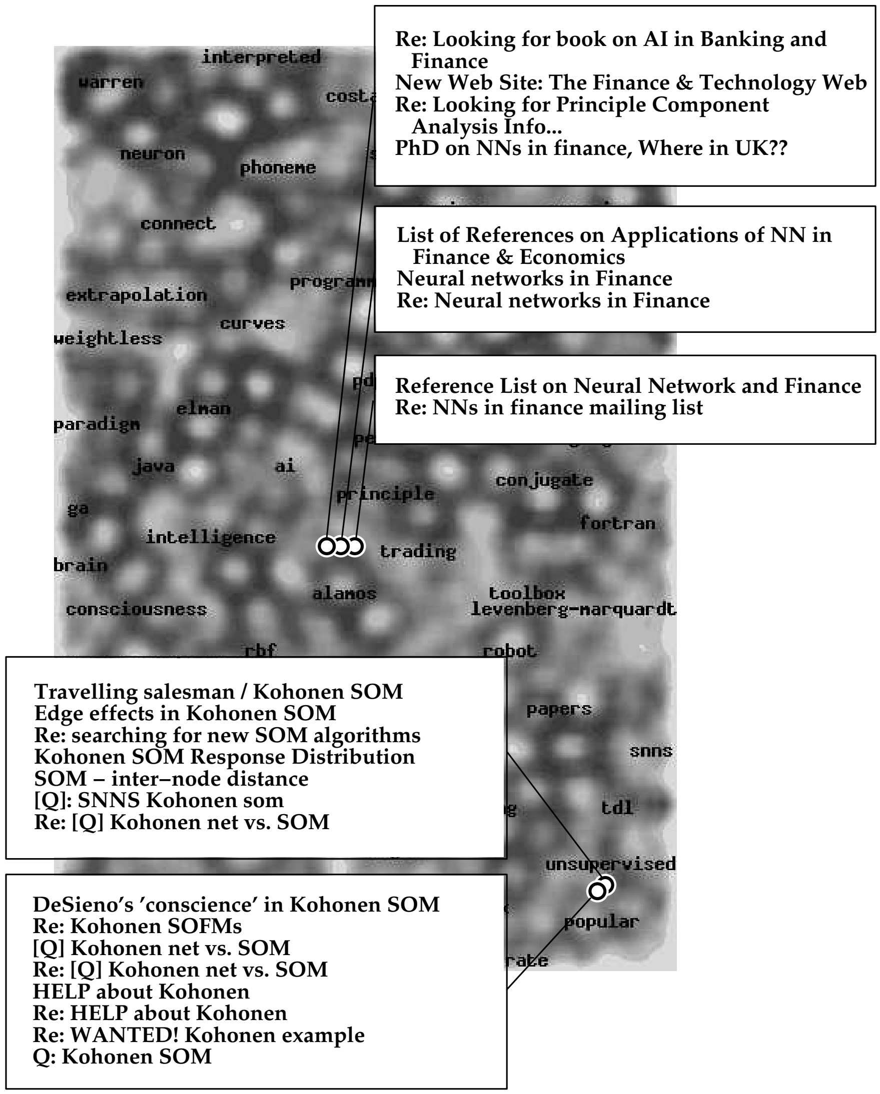


Figure 10: An example of a document map. Approximately 12000 articles that have appeared in the Usenet newsgroup comp.ai.neural-nets have been positioned on the map. The titles of the articles in five map nodes have been shown. Duplicate titles have been removed.

- A new mode of using the document maps is based on content-addressable search. The user provides a query, a list of terms or a document. The nodes that best represent the query are then highlighted on the map display to facilitate the exploration of the map.

- Recently, a method for automatic labeling of the document map has been developed. For the exploration tasks, the labels provide information on the contents of the nodes. For small maps the labels can be added manually but when large maps are considered automatic labeling is needed.

The developments mentioned above will be reported in the forthcoming publications of the WEBSOM team.

# 5 CONCLUSIONS AND FUTURE DIRECTIONS

Many issues outlined in the introduction (Chapter 1) have been worked out in the main contents of this thesis. However, some topics have been given less emphasis than what could be expected, for example the resolution of ambiguity. The introduction outlined a large research field which cannot, of course, be wholly covered in a single dissertation. The main results may be characterized to be, first, the refinement of the methods of how to use the SOM in natural language processing applications and modeling in cognitive linguistics, and, second, the development of the WEBSOM method in which the author's main contribution has been the basic design of the two-level architecture.

The number of potential future directions is large, ranging from studies of the implications of the present work in, for instance, linguistics and philosophy of language. On the other hand, further development of applications warrants investigations. Some specific possibilities are mentioned in the following.

It would be interesting to study further the relationship between modeling natural language interpretation using the SOM and some related theories. Interesting viewpoints can be found in linguistics, cognitive linguistics, cognitive science, and philosophy of language. Specific related theories include, e.g., prototype theory, cognitive grammar, relevance theory, and constructivism. The use of the SOM in natural language processing could be compared with the wide range of other statistical and corpus-based approaches.

There exist several alternatives of encoding contextual information in the texts for the word category map. Only a small sample of these was covered in this work. Moreover, other means of collecting contextual information could be developed. They may be based on, e.g., associating words with input of other modalities. New applications in areas such as machine translation and collaborative systems could be developed. In machine translation the word category maps could be used, e.g., in lexical selection, i.e., finding a suitable word or phrase to be used in a context. Text materials in languages other than English could be processed. In the current work, the various word forms of the lexemes were encoded separately. This approach seemed to work well when English is considered. If the basic form of a word may occur in over 10 000 different inflectional word forms (like verbs in Finnish) novel problems must be solved.

In the WEBSOM method the relationship between the input data characteristics, the choices of the values for the parameters, and possible preprocessing tools requires in-depth studies. As a concrete example one may consider the effect of the document length on the WEBSOM process. In some cases a long document might be divided into shorter parts which would probably have clearer identity. In traditional keyword search that uses the Boolean algebra to combine the keywords one advantage is the possibility to combine terms in various ways so that the user controls the process. Similar functionality can be achieved in the WEBSOM method by allowing the user to combine the results of subsequent searches on the document map. It would also be useful to study user satisfaction in evaluating the WEBSOM method.

# References

Alkula, R. and Honkela, T. (1992). *Development of text storage and information retrieval methods with natural language processing components. Final report of the FULLTEXT project (in Finnish).* VTT, Espoo, Finland.

Amari, S.-I. (1967). A theory of adaptive pattern classifiers. *IEEC*, 16:299–307.

Bates, M. J. (1986). Subject access in online catalog: a design model. *Journal of the American Society of Information Science*, 37(6):357–376.

Bookman, L. A. (1994). *Trajectories Through Knowledge Space: A Dynamic Framework for Machine Comprehension.* Kluwer, Boston, Dordrecht, London.

Brazdil., P. and Muggleton, S. (1991). Learning to relate terms in a multiple agent environment. In *Proceedings of Machine Learning - EWSL-91,*, pages 424–439, Berlin. Springer-Verlag.

Caramazza, A., Hillis, A., Leek, E. C., and Miozzo, M. (1994). *Mapping the mind: Domain specificity in cognition and culture*, chapter The organization of lexical knowledge in the brain: Evidence from category- and modality-specific deficits, pages 68–84. Cambridge University Press, Cambridge.

Carpenter, G. and Grossberg, S. (1987). A massively parallel architecture for a self-organizing neural pattern recognition machine. *Computer Vision, Graphics, and Image Processing*, 37:54–115.

Charniak, E. (1993). *Statistical Language Learning.* MIT Press, Cambridge, MA.

Chen, H. (1994). Collaborative systems: solving the vocabulary problem. *IEEE COMPUTER*, 27(5):58–66.

Chen, H., Schatz, B. R., Yim, T., and Fye, D. (1995). Automatic thesaurus generation for and electronic community system. *Journal of the American Society for Information Science*, 46(3):175–193.

Cottrell, M. and Fort, J.-C. (1987). Étude d'un processus d'auto-organisation. *Annales de l'Institut Henri Poincaré*, 23(1):1–20. (in French).

de Sa, V. R. (1994). *Unsupervised Classification Learning from Cross-Modal Environmental Structure.* PhD thesis, University of Rochester, Department of Computer Science, Rochester, New York.

Dickerson, J. A. and Kosko, B. (1997). *Fuzzy Engineering*, chapter Ellipsoidal fuzzy systems, pages 139–173. Prentice Hall, Upper Saddle River, NJ.

Doszkocs, T. E., Reggia, J., and Lin, X. (1990). Connectionist models and information retrieval. *Annual Review of Information Science and Technology (ARIST)*, 25.

Erwin, E., Obermayer, K., and Schulten, K. (1992a). Self-organizing maps: Ordering, convergence properties and energy functions. *Biological Cybernetics*, 67(1):47–55.

Erwin, E., Obermayer, K., and Schulten, K. (1992b). Self-organizing maps: Stationary states, metastability and convergence rate. *Biological Cybernetics*, 67(1):35–45.

Erwin, E., Obermeyer, K., and Schulten, K. (1991). Convergence properties of self-organizing maps. In Kohonen, T., Mäkisara, K., Simula, O., and Kangas, J., editors, *Artificial Neural Networks*, pages 409–414, Amsterdam, Netherlands. Elsevier.

Eysenck, M. W. and Keane, M. T. (1990). *Cognitive Psychology.* Lawrence Erlbaum Associates, Hove and London.

Faisal, K. A. and Kwasny, S. C. (1990). Design of a hybrid deterministic parser. In *Proceedings of the 13th International Conference on Computational Linguistics*, pages 11–16.

Fayyad, Piatetsky-Shapiro, Smyth, and (eds.), U. (1995). *Advances in Knowledge Discovery and Data Mining.* AAAI/MIT Press.

Flanagan, J. A. (1997). Self-organisation in the one-dimensional SOM with a reduced width neighbourhood. In *Proceedings of WSOM'97, Workshop on Self-Organizing Maps*, pages 268–273. Helsinki University of Technology, Neural Networks Research Centre, Espoo, Finland.

Fodor, J. A. and Pylyshyn, Z. W. (1988). *Connections and Symbols*, chapter Connectionism and cognitive architecture: A critical analysis. MIT Press, Cambridge, MA.

Furnas, G. W., Landauer, T. K., Gomez, L. M., and Dumais, S. T. (1987). The vocabulary problem in human-system communication. *Communications of the ACM*, 30(11):964–971.

Gallant, S. I. (1991). A practical approach for representing context and for performing word sense disambiguation using neural networks. *ACM SIGIR Forum*, 3(3):293–309.

Gärdenfors, P. (1996a). Mental representation, conceptual spaces and metaphors. *Synthese*, 106:21–47.

Gärdenfors, P. (1996b). *Philosophy and Cognitive Science*, chapter Conceptual spaces as a framework for cognitive semantics, pages 159–180. Kluwer, Dordrecht.

Haack, S. (1996). *Deviant Logic, Fuzzy Logic; Beyond the Formalism.* University of Chicago press, Chicago.

Hillis, A. E. and Caramazza, A. (1991). Category-specific naming and comprehension impairment: A double dissociation. *Brain*, 114:2081–2094.

Hirst, G. (1987). *Semantic Interpretation and the Resolution of Ambiguity.* Cambridge University Press.

Honkala, M. (1997). Personalisation through observations of browsing behaviour. Master's thesis, Helsinki University of Technology, Laboratory of Computer Science.

Honkela, T. (1989). Semantic interpretation and knowledge acquisition in natural language database interface (in Finnish). Master's thesis, University of Oulu, Department of Information Processing Science.

Honkela, T., Kaski, S., Lagus, K., and Kohonen, T. (1996). Newsgroup exploration with WEBSOM method and browsing interface. Technical Report A32, Helsinki University of Technology, Laboratory of Computer and Information Science, Espoo, Finland.

Hörmann, H. (1986). *Meaning and Context.* Plenum Press, New York.

Horowitz, R. and Alvarez, L. (1996). Self-organizing neural networks: convergence properties. In *Proceedings of ICNN'96, 1996 IEEE International Conference on Neural Networks*, volume 1, pages 7–12. IEEE, New York, NY, USA.

Hyötyniemi, H. (1996). Text document classification with self-organizing maps. In Alander, J., Honkela, T., and Jakobsson, M., editors, *Proceedings of Finnish Artificial Intelligence Conference – Genes, Nets and Symbols*, pages 64–72. Finnish Artificial Intelligence Society.

Hyötyniemi, H. and Lehtola, A. (1988). UQL – the universal relation query language. In *Proceedings of STeP-88, Finnish Artificial Intelligence Symposium*, pages 803–810. Finnish Artificial Intelligence Society.

Hyötyniemi, H. and Lehtola, A. (1991). A universal relation database interface for knowledge based systems. In *Proceedings of DASFAA'91, 2nd International Symposium on Database Systems for Advanced Applications*, pages 84–88. Tokyo Information Processing Society of Japan.

Jäppinen, H., Honkela, T., Hyötyniemi, H., and Lehtola, A. (1988a). Hierarchical multilevel processing model for natural language database interface. In *Proceedings of CAIA'88, 4th Conference on Artificial Intelligence Applications*, pages 332–337. IEEE (The Computer Society of the IEEE).

Jäppinen, H., Honkela, T., Hyötyniemi, H., and Lehtola, A. (1988b). A multilevel natural language processing model. *Nordic Journal of Linguistics*, 11:69–82.

Jäppinen, H., Lehtola, A., Nelimarkka, E., and Ylilammi, M. (1983). Knowledge engineering approach to morphological analysis. In *Proceedings of the First Conference of the European Chapter of the Association for Computational Linguistics*, pages 49–51.

Jäppinen, H. and Ylilammi, M. (1986). Associative model of morphological analysis: An empirical inquiry. *Computational Linguistics*, 12(4).

Järvinen, T. (1994). Annotating 200 million words: The bank of English project. In *Proceedings of Coling-94, Vol. I*, pages 565–568.

Karlsson, F. (1990). Constraint grammar as a framework for parsing running text. In Karlgren, H., editor, *Papers presented to the 13th International Conference on Computational Linguistics, volume 3*, pages 168–173. International Committee on Computational Linguistics.

Karlsson, F., Voutilainen, A., Heikkilä, J., and Anttila, A. (1995a). *Constraint Grammar: A Language-Independent Framework for Parsing Unrestricted Text*, chapter Designing a Parser for Unrestricted Text, pages 1–40. Mouton de Gruyter, Berlin / New York.

Karlsson, F., Voutilainen, A., Heikkilä, J., and Anttila, A. (1995b). *Constraint Grammar: A Language-Independent Framework for Parsing Unrestricted Text*, chapter The Formalism and Environment of Constraint Grammar Parsing, pages 41–88. Mouton de Gruyter, Berlin / New York.

Karlsson, F., Voutilainen, A., Heikkilä, J., and (eds.), A. A. (1995c). *Constraint Grammar: A Language-Independent Framework for Parsing Unrestricted Text*. Mouton de Gruyter, Berlin / New York.

Kaski, S. (1997a). Computationally efficient approximation of a probabilistic model for document representation in the WEBSOM full-text analysis method. *Neural Processing Letters*, 5(2):69–81.

Kaski, S. (1997b). Data exploration using self-organizing maps. *Acta Polytechnica Scandinavica, Mathematics, Computing and Management in Engineering Series No. 82*. DTech Thesis, Helsinki University of Technology, Finland.

Kaski, S. (1998). Dimensionality reduction by random mapping: Fast similarity computation for clustering. In *Proceedings of IJCNN'98, International Joint Conference on Neural Networks, submitted for publication.*

Kaski, S. and Lagus, K. (1996). Comparing self-organizing maps. In von der Malsburg, C., von Seelen, W., Vorbrüggen, J. C., and Sendhoff, B., editors, *Proceedings of ICANN96, International Conference on Artificial Neural Networks, Bochum, Germany, July 16-19*, Lecture Notes in Computer Science, vol. 1112, pages 809–814. Springer, Berlin.

Kohonen, T. (1981a). Automatic formation of topological maps of patterns in a self-organizing system. In Oja, E. and Simula, O., editors, *Proceedings of 2SCIA, Scand. Conference on Image Analysis*, pages 214–220, Helsinki, Finland. Suomen Hahmontunnistustutkimuksen Seura r.y.

Kohonen, T. (1981b). Construction of similarity diagrams for phonemes by a self-organizing algorithm. Technical Report TKK-F-A463, Helsinki University of Technology, Espoo, Finland.

Kohonen, T. (1981c). Hierarchical ordering of vectorial data in a self-organizing process. Report TKK-F-A461, Helsinki University of Technology, Espoo, Finland.

Kohonen, T. (1981d). Self-organized formation of generalized topological maps of observations in a physical system. Report TKK-F-A450, Helsinki University of Technology, Espoo, Finland.

Kohonen, T. (1982a). Analysis of a simple self-organizing process. *Biological Cybernetics*, 44(2):135–140.

Kohonen, T. (1982b). Self-organizing formation of topologically correct feature maps. *Biological Cybernetics*, 43(1):59–69.

Kohonen, T. (1988). The 'neural' phonetic typewriter. *Computer*, 21(3):11–22.

Kohonen, T. (1992). An attempt to interpret the Self-Organizing Mapping physiologically. Report A16, Helsinki University of Technology, Laboratory of Computer and Information Science.

Kohonen, T. (1993). Physiolocigal interpretation of the self-organizing map algorithm. *Neural Networks*, 6(7):895–905.

Kohonen, T. (1995a). The Adaptive-Subspace SOM (ASSOM) and its use for the implementation of invariant feature detection. In Fogelman-Soulié, F. and Gallinari, P., editors, *Proceedings of ICANN'95, International Conference on Artificial Neural Networks*, volume I, pages 3–10, Nanterre, France. EC2.

Kohonen, T. (1995b). *Computational Intelligence - A Dynamic System Perspective*, chapter 2. Emergence of Invariant-Feature Detectors in Self-Organization, pages 17–31. IEEE Press, New York.

Kohonen, T. (1995c). *Self-Organizing Maps*. Springer, Berlin, Heidelberg.

Kohonen, T. (1996a). Emergence of invariant-feature detectors in the adaptive-subspace self-organizing map. *Biological Cybernetics*, 75(4):281–91.

Kohonen, T. (1996b). The self-organizing map, a possible model of brain maps. *Medical & Biological Engineering & Computing*, 34(suppl.1, pt.1):5–8.

Kohonen, T. (1997). Exploration of very large databases by self-organizing maps. In *Proceedings of ICNN'97, International Conference on Neural Networks*, pages PL1–PL6. IEEE Service Center, Piscataway, NJ.

Kohonen, T., Hynninen, J., Kangas, J., and Laaksonen, J. (1996a). SOM_PAK: The Self-Organizing Map program package. Report A31, Helsinki University of Technology, Laboratory of Computer and Information Science.

Kohonen, T., Kaski, S., Lagus, K., and Honkela, T. (1996b). Very large two-level SOM for the browsing of newsgroups. In von der Malsburg, C., von Seelen, W., Vorbrüggen, J. C., and Sendhoff, B., editors, *Proceedings of ICANN96, International Conference on Artificial Neural Networks, Bochum, Germany, July 16-19, 1996*, Lecture Notes in Computer Science, vol. 1112, pages 269–274. Springer, Berlin.

Kohonen, T., Kaski, S., and Lappalainen, H. (1997a). Self-organized formation of various invariant-feature filters in the adaptive-subspace som. *Neural Computation*, 9:1321–1344.

Kohonen, T., Kaski, S., Lappalainen, H., and Salojärvi, J. (1997b). The adaptive-subspace self-organizing map (ASSOM). In *Proceedings of WSOM'97, Workshop on Self-Organizing Maps, Espoo, Finland, June 4-6*, pages 191–196. Helsinki University of Technology, Neural Networks Research Centre, Espoo, Finland.

Kohonen, T., Mäkisara, K., and Saramäki, T. (1984). Phonotopic maps - insightful representation of phonological features for speech recognition. In *Proceedings of 7ICPR, International Conference on Pattern Recognition*, pages 182–185, Los Alamitos, CA. IEEE Computer Soc. Press.

Kohonen, T., Oja, E., Simula, O., Visa, A., and Kangas, J. (1996c). Engineering applications of the self-organizing map. *Proceedings of the IEEE*, 84(10):1358–84.

Koikkalainen, P. (1994). Progress with the tree-structured self-organizing map. In Cohn, A. G., editor, *Proceedings of ECAI'94, 11th European Conference on Artificial Intelligence*, pages 211–215, New York. John Wiley & Sons.

Koikkalainen, P. and Oja, E. (1990). Self-organizing hierarchical feature maps. In *Proc. IJCNN-90-Wash-DC, International Joint Conference on Neural Networks*, volume II, pages 279–285, Piscataway, NJ. IEEE Service Center.

Koskenniemi, K. (1983). *Two-Level Morphology: A General Computational Model for Word-Form Recognition and Production*. PhD thesis, University of Helsinki, Department of General Linguistics.

Kosko, B. (1997). *Fuzzy Engineering*. Prentice Hall, Upper Saddle River, NJ.

Lagus, K. (1997). Map of WSOM'97 abstracts—alternative index. In *Proceedings of WSOM'97, Workshop on Self-Organizing Maps, Espoo, Finland, June 4-6*, pages 368–372. Helsinki University of Technology, Neural Networks Research Centre, Espoo, Finland.

Lagus, K., Honkela, T., Kaski, S., and Kohonen, T. (1996a). Self-organizing maps of document collections: A new approach to interactive exploration. In Simoudis, E., Han, J., and Fayyad, U., editors, *Proceedings of the Second International Conference on Knowledge Discovery and Data Mining*, pages 238–243. AAAI Press, Menlo Park, California.

Lagus, K., Honkela, T., Kaski, S., and Kohonen, T. (1996b). WEBSOM – a status report. In Alander, J., Honkela, T., and Jakobsson, M., editors, *Proceedings of STeP'96, Finnish Artificial Intelligence Conference*, pages 73–78. Finnish Artificial Intelligence Society, Vaasa, Finland.

Lagus, K., Kaski, S., Honkela, T., and Kohonen, T. (1996c). Browsing digital libraries with the aid of self-organizing maps. In *Proceedings of the Fifth International World Wide Web Conference WWW5, May 6-10, Paris, France*, volume Poster Proceedings, pages 71–79. EPGL.

Lakoff, G. (1987). *Women, Fire and Dangerous Things*. University of Chicago Press, Chicago.

Lehtola, A. and Honkela, T. (1988). AWARE - a transformation system for semantic analysis of natural language. In *Proceedings of STeP-88, Finnish Artificial Intelligence Symposium*, pages 100–109. Finnish Artificial Intelligence Society.

Lehtola, A., Honkela, T., Hyötyniemi, H., and Jäppinen, H. (1988a). Knowledge languages and metatools for natural language processing. In *Pro-*

*ceedings of ICSC'88, International Computer Science Conference'88.*, pages 273–280. IEEE Computer Society, IEEE Hongkong Section.

Lehtola, A., Honkela, T., Hyötyniemi, H., and Jäppinen, H. (1988b). Task oriented knowledge representation languages for nlp-systems. In *Proceedings of ISMIS'88, 3rd International Symposium on Methodologies for Intelligent Systems*, pages 250–259. North-Holland.

Lehtola, A. and Valkonen, K. (1986). Efficient search using blackboards - a view into Finnish parsing. In Karjalainen, editor, *Proceedings of the Third Finnish Symposium on Theoretical Computer Science*, pages 64–87. University of Joensuu.

Leivian, R., Peterson, W., and Gardner, M. (1997). CorDex: a knowledge discovery tool. In *Proceedings of WSOM'97, Workshop on Self-Organizing Maps*, pages 63–68. Helsinki University of Technology, Neural Networks Research Centre.

Lenat, D. B., Guha, R. V., Pittman, K., Pratt, D., and Shepherd, M. (1990). Cyc: Toward programs with common sense. *Communications of the ACM*, 33:30–49.

Levinson, S. (1983). *Pragmatics.* Cambridge University Press, Cambridge.

Lin, X. (1992). Visualization for the document space. In *Proceedings of Visualization '92*, pages 274–81, Los Alamitos, CA, USA. Center for Comput. Legal Res., Pace Univ., White Plains, NY, USA, IEEE Comput. Soc. Press.

Lin, X. (1997). Map displays for information retrieval. *Journal of the American Society for Information Science*, 48:40–54.

Lin, X., Soergel, D., and Marchionini, G. (1991). A self-organizing semantic map for information retrieval. In *Proceedings of 14th. Ann. International ACM/SIGIR Conference on Research & Development in Information Retrieval*, pages 262–269.

Lindén, K. (1993). Finnish language technology. In *Life and Education in Finland.*

Linell, P. (1982). *The Written Language Bias in Linguistics.* Linköping.

MacLennan, B. (1995). Continuous formal systems: A unifying model in language and cognition. In *Proceedings of the IEEE Workshop on Architectures for Semiotic Modeling and Situation Analysis in Large Complex Systems, August 27-29, 1995, Monterey, CA.*

MacWhinney, B. (1989). *Linguistic categorization*, chapter Competition and Lexical Categorization. Benjamins, New York.

MacWhinney, B. (1997). *Cognitive approaches to language learning*, chapter Lexical Connectionism. MIT Press.

Mayberry III, M. R. and Miikkulainen, R. (1994). Lexical disambiguation based on distributed representations of context frequency. In *Proceedings of the 16th Annual Conference of the Cognitive Science Society*.

McKenna, P. and Warrington, E. K. (1978). Category-specific naming preservation: A single case study. *Journal of Neurology, NeuroSurgery, and Psychiatry*, 41:571–574.

Merkl, D. (1993). Structuring software for reuse - the case of self-organizing maps. In *Proceedings of IJCNN-93-Nagoya, International Joint Conference on Neural Networks*, volume III, pages 2468–2471, Piscataway, NJ. IEEE Service Center.

Merkl, D. (1994). *Self-Organization of Software Libraries: An Artificial Neural Network Approach*. PhD thesis, Institut für Angewandte Informatik und Informationssysteme, Universität Wien.

Merkl, D. (1995a). Content-based document classification with highly compressed input data. In Fogelman-Soulié, F. and Gallinari, P., editors, *Proceedings of ICANN'95, International Conference on Artificial Neural Networks*, volume II, pages 239–244, Nanterre, France. EC2.

Merkl, D. (1995b). Content-based software classification by self-organization. In *Proceedings of ICNN'95, IEEE International Conference on Neural Networks*, volume II, pages 1086–1091, Piscataway, NJ. IEEE Service Center.

Merkl, D. (1997). Lessons learned in text document classification. In *Proceedings of WSOM'97, Workshop on Self-Organizing Maps, Espoo, Finland, June 4-6*, pages 316–321. Helsinki University of Technology, Neural Networks Research Centre, Espoo, Finland.

Merkl, D. and Rauber, A. (1997). Alternative ways for cluster visualization in self-organizing maps. In *Proceedings of WSOM'97, Workshop on Self-Organizing Maps, Espoo, Finland, June 4-6*, pages 106–111. Helsinki University of Technology, Neural Networks Research Centre, Espoo, Finland.

Merkl, D., Tjoa, A. M., and Kappel, G. (1994). Application of self-organizing feature maps with lateral inhibition to structure a library of reusable sotware components. In *Proceedings of ICNN'94, International Conference on Neural Networks*, pages 3905–3908, Piscataway, NJ. IEEE Service Center.

Miikkulainen, R. (1990a). *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).

Miikkulainen, R. (1990b). A distributed feature map model of the lexicon. In *Proceedings of 12th Annual Conference of the Cognitive Science Society*, pages 447–454, Hillsdale, NJ. Lawrence Erlbaum.

Miikkulainen, R. (1991). A neural network model of script processing and memory. In *Proceedings of International Workshop on Fundamental Res. for the Next Generation of Natural Language Processing*, Kyoto, Japan. ATR International.

Miikkulainen, R. (1993a). DISCERN: A distributed neural network model of script processing and memory. In *Proceedings of Third Twente Workshop on Language Technology*, Twente, Netherlands. Computer Science Department, University of Twente.

Miikkulainen, R. (1993b). *Subsymbolic Natural Language Processing: An Integrated Model of Scripts, Lexicon, and Memory.* MIT Press, Cambridge, MA.

Miikkulainen, R. (1997). Self-organizing feature map model of the lexicon. *Brain and Language*, 59:334–366.

Miikkulainen, R. and Dyer, M. G. (1991). Natural language processing with modular neural networks and distributed lexicon. *Cognitive Science*, 15:343–399.

Mitra, S. and Pal, S. (1994). Self-organizing neural network as a fuzzy classifier. *IEEE Transactions on Systems, Man and Cybernetics*, 24(3):385–99.

Mitra, S. and Pal, S. (1996). Fuzzy self-organization, inferencing, and rule generation. *IEEE Transactions on Systems, Man & Cybernetics, Part A [Systems & Humans]*, 26(5):608–20.

Moore, T. and Carling, C. (1988). *The Limitations of Language*. Macmillan Press, Houndmills.

Nakamura, M., Maruyama, K., Kawabata, T., and Shikano, K. (1990). Neural network for word category prediction. In Karlgren, H., editor, *Papers presented to the 13th International Conference On Computational Linguistics.* International Committee on Computational Linguistics.

Nenov, V. I. and Dyer, M. G. (1993). Perceptually grounded language learning: Part 1 - A neural network architecture for robust sequence association. *Connection Science*, 5(2):115–138.

Nenov, V. I. and Dyer, M. G. (1994). Perceptually grounded language learning: Part 2 - DETE: a neural/procedural model. *Connection Science*, 6(1):3–41.

Persson, I.-B. (1995). *Connectionism, Language Production and Adult Aphasia; Elaboration of a Connectionist Framework for Lexical Processing and a Hypothesis of Agrammatic Aphasia.* PhD thesis, University of Helsinki, Societas Scientiarum Fennica.

Pollack, J. B. (1990). Recursive distributed representations. *Artificial Intelligence*, 46:77–105.

Pulkki, V. (1995). Data averaging inside categories with the self-organizing map. Report A27, Helsinki University of Technology, Laboratory of Computer and Information Science, Espoo, Finland.

Raukko, J. (1994). Polysemy: network of meanings - net of meaning (in Finnish). In Leino, P. and Onikki, T., editors, *Perspectives on polysemy. Kieli 8. Cognitive Grammar on Finnish 2. (in Finnish)*, pages 36–69. Department of Finnish, University of Helsinki, Helsinki.

Raukko, J. (1997). *Cultural, Typological, and Psychological Perspectives in Cognitive Linguistics*, chapter An 'intersubjective' method for cognitive-semantic research on polysemy: The case of _get_.

Regier, T. (1995). A model of the human capacity for categorizing spatial relations. *Cognitive Linguistics*, 6(1):63–88.

Ritter, H. and Kohonen, T. (1989). Self-organizing semantic maps. *Biological Cybernetics*, 61(4):241–254.

Ritter, H. and Kohonen, T. (1990). Learning 'semantotopic maps' from context. In *Proceedings of IJCNN-90-WASH-DC, International Joint Conference on Neural Networks*, volume I, pages 23–26, Hillsdale, NJ. Lawrence Erlbaum.

Rosch, E. (1977). *Studies in cross-cultural psychology, vol. 1*, chapter Human categorization, pages 3–49. Academic Press, New York.

Rozmus, J. (1995). Information retrieval by self-organizing maps. In Williams, M., editor, *16th National Online Meeting Proceedings - 1995*, pages 349–54, Medford, NJ, USA. Smart Syst., USA, Learned Inf.

Rumelhart, D. E. and McClelland, J. L. (1986). *Parallel Distributed Processing; Volume 2: Psychological and Biological Models*, chapter On learning the past tenses of English verbs, pages 216–271. MIT Press, Cambridge, MA.

Salton, G. (1989). *Automatic Text Processing.* Addison-Wesley, Reading, MA.

Salton, G., Wong, A., and Yang, C. S. (1975). A vector space model for automatic indexing. *Communications of the ACM*, 18(11):613–620.

Schnelle, H. and Wilkens, R. (1990). Connectionist networks for constituent structure. In Karlgren, H., editor, *Papers presented to the 13th International Conference On Computational Linguistics.* International Committee on Computational Linguistics.

Scholtes, J. C. (1992a). Neural data oriented parsing. In *Proceedings of 3rd Twente Workshop on Language Technology*, Twente, Netherlands. University of Twente.

Scholtes, J. C. (1992b). Resolving linguistic ambiguities with a neural data-oriented parsing (DOP) system. In Aleksander, I. and Taylor, J., editors, *Artificial Neural Networks, 2*, volume II, pages 1347–1350, Amsterdam, Netherlands. North-Holland.

Scholtes, J. C. (1993). *Neural Networks in Natural Language Processing and Information Retrieval.* PhD thesis, Universiteit van Amsterdam, Amsterdam, Netherlands.

Scholtes, J. C. and Bloembergen, S. (1992). The design of a neural data-oriented parsing (DOP) model. In *Proceedings of IJCNN-92-Baltimore, International Joint Conference on Neural Networks*, volume II, pages 69–72, Piscataway, NJ. IEEE Service Center.

Schütze, H. (1992). Dimensions of meaning. In *Proceedings of Supercomputing*, pages 787–796.

Tikkala, A., Eikmeyer, H.-J., Niemi, J., and Laine, M. (1997). On the production of Finnish nouns: A psycholinguistically motivated connectionist model. *Connection Science, in press.*

Tirri, H. (1991). Implementing expert system rule conditions by neural networks. *New Generation Computing*, 10(1):55–71.

Ultsch, A. (1993). Self-organizing neural networks for visualization and classification. In Opitz, O., Lausen, B., and Klar, R., editors, *Information and Classification*, pages 307–313, London, UK. Springer.

Ultsch, A. and Siemon, H. (1990). Kohonen's self organizing feature maps for exploratory data analysis. In *Proceedings of INNC'90, International Neural Network Conference*, pages 305–308, Dordrecht, Netherlands. Kluwer.

Valkonen, K., Jäppinen, H., and Lehtola, A. (1987). Blackboard-based dependency parsing. In *Proceedings of IJCAI'87, Tenth International Joint Conference on Artificial Intelligence*, pages 700–702.

Varela, F. J., Thompson, E., , and Rosch, E. (1993). *The Embodied Mind: Cognitive Science and Human Experience*. MIT Press, Cambridge, Massachusetts.

von der Malsburg, C. (1973). Self-organization of orientation sensitive cells in the striate cortex. *Kybernetik*, 14:85–100.

Wermter, S., Riloff, E., and Scheler, G. (1996). *Connectionist, Statistical and Symbolic Approaches to Learning for Natural Language Processing*. Springer Verlag, New York.

Winston, P. H. (1984). *Artificial Intelligence, Second Edition*. Addison-Wesley, Reading, MA.

Zadeh, L. A. (1965). Fuzzy sets. *Information and Control*, 8:338–353.

Zavrel, J. (1995). Neural information retrieval - an experimental study of clustering and browsing of document collections with neural networks. Master's thesis, University of Amsterdam, Amsterdam, Netherlands.

Zavrel, J. (1996). Neural navigation interfaces for information retrieval: are they more than an appealing idea? *Artificial Intelligence Review*, 10(5-6):477–504.