Neural Network Models for Language Acquisition: A Brief Survey

Jordi Poveda[1] and Alfredo Vellido[2]

[1] TALP Research Center
[2] Soft Computing Research Group
Technical University of Catalonia (UPC), Barcelona, Spain
{jpoveda,avellido}@lsi.upc.edu

Abstract. Since the outbreak of connectionist modelling in the mid eighties, several problems in natural language processing have been tackled by employing neural network-based techniques. Neural network's "biological plausibility" offers a promising framework in which the computational treatment of language may be linked to other disciplines such as cognitive science and psychology. With this brief survey, we set out to explore the landscape of artificial neural models for the acquisition of language that have been proposed in the research literature.

1 Introduction

Human language, as a canonical representative of human cognitive faculties, has gathered wide attention in such diverse fields as cognitive science, psychology, artificial intelligence and, of course, linguistics. There are strong intuitive reasons to believe that human cognition, at least at its higher levels, revolves around mental representations that have language at the base. This is why a better understanding of the mechanisms behind language acquisition and its representation in the brain could shed some light in unresolved questions about the working of the human mind. In this respect, the ability to have computional models run and interact with linguistic input data, and to analyze quantitave and qualitative results, plays a very important role. Whether artificial neural networks (ANNs) provide meaningful models of the brain and to what degree, and whether they constitute a useful approach to natural language processing (NLP) is subject to debate [1]. Throughout this survey, we will examine some of the main arguments raised for and against the explanatory potential of ANNs as models for language acquisition, while supporting the position that they do indeed possess at least potential as useful tools and models.

First language acquisition concerns itself with the processes involved in the development of language in children. There have been traditionally two schools of thought: nativists and non-nativists or "emergencionists". Nativists assume that the ability for language is for the most part innate, and thus the underlying principles of language are universal and inborn to all humans. Proponents of nativism are Chomsky, Fodor and Pinker, among others. A central idea to nativism is the well-known Chomskian postulate of the existence of a Universal Grammar [2], which is innate to all individuals and which underlies all specific instances of human languages. Non nativists (among them, Mac Whinney, Bates and Snow), despite admitting that some of the ability for developing language may be innate, see language acquisition as a rather emergent process and a result of children's social interaction and exposure to linguistic stimuli. As we will see later, connectionist models of the brain in general, and self-organizing models in particular, due to their design characteristics, have a lot to say as advocates of the view of language acquisition as an emergent process.

The acquisition of a second language and bilingual development in children present their own of set of issues: the interference and coupling effects between the two languages, how the two languages and their respective lexicons are represented in the brain, the effects of age of acquisition, etc. We will also see a neural-network based model of bilingualism that particially addresses these issues.

The goal of this survey is to explore ANN-based approaches for an specific NLP problem: that of language acquisition; what research efforts have historically been made, where this area of research currently stands, and to what degree ANNs are viable and biologically plausible models of language acquisition. It was conceived in the light of a perceived prevalence of statistical (e.g. HMMs, linear classifiers, Gaussian models, SVMs, ...) and relational (rule induction) methods for NLP problems in general in current AI research, in detriment of more generalized use of ANN (connectionist) methods, although connectionist models of language cognitive development are being reappraised [3]. Nonetheless, it can also be argued that statistical and connectionist methods are not necessarily mutually exclusive fields. Much work has been done both to provide a probabilistic interpretation of neural networks and to insert neural networks within a probabilistic framework [4,5,6]. Some of the most recent theories of cortical activity draw heavily both from connectionist models and from probability theory [7].

The rest of this work is structured as follows. In Sect. 2, we provide some historical perspective on connectionist modelling of natural language. Section 3 deals with the strengths of two specific neural architectures that have been proposed to model aspects of language acquisition, namely Self-Organizing Maps (SOM) and the Simple Recurrent Network (SRN). In Sect. 4, we describe four particular proposed architectures commented in a greater level of detail: the ANN for learning english verbs past tenses by Rummelhart and McClelland [8], TRACE [9], SOMBIP [10] and DevLex [11]. We will examine in turn the rationale behind these models, their architectures, training methods and their main results and implications. Section 5 briefly compares connectionist modelling of lexical acquisition with statistical and other approaches. Finally, Sect. 6 presents the conclusions.

2 The Connectionism vs. Symbolism Controversy: An Historical Perspective

Connectionism attempts to construct biologically-inspired computacional models of cognitive processes as a network of connections between processing units or nodes, known as neurons [12]. According to this view of computation, information is stored in the form of weights between the nodes' connections, or synapses, in imitation of biological synapses. Connectionist models of NLP took off in the late eighties thanks to the pioneering work of Rummelhart & McClelland [8], with their famous ANN model of the acquisition of past tenses of English verbs. Rummelhart and McClelland's model was intended as a proof-of-concept against symbolism (and simultaneously against the prevailing nativist view of the time that language ability was hardwired into the brain from birth): they argued, information could be better captured in the form of connections among processing units, thus eliminating the need for formulating explicit rules that try to explain the details of acquisition phenomena. This claim was widely contested from symbolist circles (e.g. [13]).

The proponents of symbolism picture the human brain as a digital processor of symbolic information, and argue that computational models of the brain should be based on algorithmic programs manipulating symbols. This is the traditional school of thought, antagonist to connectionism, which denies the validity of connectionist models altogether and doesn't credit them with any explanatory potential. A halfway position between these two opposite views is that of implementational connectionists (e.g. Fodor, Pinker and Pylyshyn), who admit the utility of ANNs in modelling cognitive processes, but hold that they should be employed ultimately to implement symbolic processing ("*the mind is a neural net; but it is also a symbolic processor at a higher and more abstract level of description*" [14]). According to them, research of models should be made at the symbolic (psychological) level, whereas ANNs are the tools through which these models are implemented in practice.

Fodor raised in [15] a well-known argument against the adequacy of connectionism as a model of the mind, based on a characteristic of human intelligence which he called systematicity. Neural networks, he said, are good at capturing associations, but they alone cannot account for higher cognitive abilities required, for instance, for human language. Still another main criticism against connectionist models of language is based on the *compositionality* of language (the meaning of a complex statement can be decomposed in terms of the individual meanings of its simpler constituents). As if to contest this challenge launched against connectionism about the recursive nature of language, Pollack devised a neural network architecture that was well-suited to represent recursive data structures, such as trees and lists: the recursive auto-associative memory (RAAM) [16]. Due to their ability to represent recursive data structures, RAAM networks are useful for working with syntactic and semantic representations in NLP applications. In the field of speech processing, the TRACE architecture by McClelland and Elman [9] set another milestone in early connectionist modelling of language.

More recent systems have used SOM as neural-network models of language acquisition. One such model is Miikkulainen's DISLEX [17], which is composed of multiple self-organizing feature maps. DISLEX is a neural network model of the mental lexicon, intented to explain the phenomena involved in lexical aphasia. We will conclude this historical revision by making a reference to the CHILDES database project [18]. The CHILDES database is a corpus of child-directed speech, that is, recordings and transcripts of conversations between parents and young children. It has been subsequently used by other experiments on neural network modelling of lexical acquisition, in order to gather training data for the model, and to build a restricted lexicon, representative of the first stages of language learning.

3 Artificial Neural Network Architectures for Language Acquisition

In this section we will discuss how two specific neural network architectures, Kohonen's SOM [19] and Elman's SRN [20], have been applied for modelling aspects of language acquisition and have served as building blocks for larger ANN models.

3.1 Kohonen Self-Organizing Maps

A SOM network defines a topology-preserving mapping between a often highly dimensional input space and a low dimensional, most typically 2-D, space. Self-organization is introduced by having the notion of neighbouring units, whose weights are adjusted in proportion to their distance from the winning unit. Several characteristics of SOM make this architecture especially suitable for modelling language acquisition [10]:

1. *Unsupervised learning*: SOM is trained by presenting inputs to the network (without correcting feedback). This is coherent with the way in which children are for the most part exposed to language.
2. *Self-organization*: Activation of the best-matching unit and propagation of activation yield network units that specialize in specific groups of related words, and resonance between the input and the matching neuron(s) is increased. This presents a coherent picture of memory and the process of remembering.
3. *Representation*: Inputs that are close in the high dimensional space will activate nearby units in the 2-D space. Also, semantic categories emerge in SOM in the form of clusters of related words.
4. *Neighbourhood function*: Acting on the neighbourhood ratio allows the modelling of different levels of brain plasticity. Early plasticity and formation of gross categories, and posterior establishment and fine-grained specialization of the learned structures can be modelled by decreasing the neighbourhood ratio through the learning process.

5. *Hebbian learning*: SOM maps interconnected through hebbian associative links can be used to model the interactions among different levels of language, as is done in DISLEX architecture [17].

Anderson reports in [21] the results of several experiments he conducted with SOM simulations in order to model a number of aspects of language acquisition, including: the modelization of the process of learning to distinguish word boundaries in a continuous stream of speech; the modelization of the disappearance with age of the ability to recognize phonemes other than those of one's own language; and the modelization of the clinical occurrences of semantically bounded anomia (i.e. inability to distinguish correctly among words belonging to some semantic category).

3.2 SRN for Building Word Meaning Representations

A simple recurrent network (SRN) architecture, as introduced by Elman in [20], can be employed to construct distributed representations (i.e. as a vector of weights) for the meaning of a word. The word meaning representations are built from contextual features, by putting the word in relation to its context, as it occurs in a stream of input sentences. This is indeed what Li and Farkas do in the WCD (Word Co-ocurrence Detector) subsystem of their DevLex [11] and SOMBIP [10] models, both of which are described in Sect. 4.

The SRN network has two layers, an input layer and a hidden layer (which we will call copy layer). This model assigns to each word w_i of a lexicon of size N a unary encoding as a vector of N dimensions, where the i-th component is 1 and the rest of components are 0. The input layer has N input units (as many units as the number of components in a word's encoding). At each time instant t, the hidden or copy layer contains a one-to-one copy of the previous vector on the input layer (the input word at time $t-1$). L and R are two arrays of associative vectors, fully connecting the units of the copy layer to input layer and viceversa. Training consists in presenting the network with words from a stream of input sentences, one word at a time. The weight l_{ij}, connecting unit j in the copy layer to unit i in the input layer, expresses the probability $P(j^{t-1}|i^t)$ that word w_i is preceded by word w_j. Similarly, the weight r_{ij} expresses the probability $P(i^t|j^{t-1})$ that word w_i follows word w_j. These weights are updated by hebbian learning after each input word is presented. By the end of the training, $l_i = [l_{i1} \dots l_{iN}]$ contains a representation of the left context of word w_i (the probability distribution of the words preceding i), and $r_i = [r_{1i} \dots r_{Ni}]$ contains a representation of the right context of w_i (the probability distribution of the words following i). The concatenation of these two vectors forms the distributed representation of the meaning of a word.

4 Case Studies

In the previous sections, we have examined general questions about the subject of language acquisition, trying to relate the viewpoints of different disciplines.

In this section we examine four complete ANN models that have been proposed in the literature to tackle different problems in language acquisition. These four models are presented here in chronological order of appearance in the research literature, so that the reader will realize how each one draws on the experience and foundations laid out by the previous ones.

A first attempt to establish a typology of neural lexical models may be established with regard to the type of representation they use, and to the behaviour of the network over time: in *localist* representations, each word or the meaning of the concept that it conveys is represented by a single neuron or processing unit (i.e. localized), whereas in *distributed* representations, the representation of each word or its corresponding concept is spread through multiple units of the network; likewise, regarding evolution with time, *stationary* or permanent models are those in which the connection weights (and the network architecture) are prespecified, whereas in *dynamic* or learning models the connection weights (and/or the network architecture) evolve through time. The TRACE architecture, for instance, is a localist and stationary model. In constrast, systems based on the SRN architecture introduced by Elman [20] are usually distributed and dynamic. Dynamic models afford a better interpretability of the observed results than stationary ones, by putting the model dynamics in relation with the dynamics of human lexical learning evidenced by psychology and cognitive science experimentation. Localist and distributed representations serve different purposes and are not mutually exclusive: some complex multi-level ANN models such as DevLex and others based on SOM maps exhibit both types of unit-word correspondence simultaneously at different levels of representation.

A second attempt at establishing a taxonomy of these models refers to the type of basic ANN architecture underlying the model. Table 1 summarizes this distinction and presents some highlights of each type of model.

4.1 Rummelhart and McClelland: Acquisition of Past Tenses of English Verbs

Rummelhart and McClelland [8] used a one-layer feedforward network based on the perceptron learning algorithm in order to map verb roots to their past tense forms. The representation of verbs was based on a system of phonological features (Wickelphones), into which verb roots were encoded prior to being inputs to the network, and which were decoded at the output. Rummelhart and McClelland wanted to model the U-shaped learning curve typically found in children: early correct production of a few irregular verbs, middle confusion due to mixing of regular and irregular verbs' patterns, and late correct production of the majority of verbs. To this end, they split a training of 200 epochs in two stages: in the first 10 epochs, they presented the network with 10 highly-ocurring verbs; later, during the remaining 190 epochs, they introduced 410 medium-frequency verbs. The testing set consisted of 86 low-frequency verbs (14 irregular and 72 regular). They report having observed the U-shaped pattern of learning, as many irregulars were incorrectly produced during the middle stages of training due to overregularization.

Table 1. A comparison of ANN models of lexical acquisiton by underlying architecture

Type	Examples	Training	Highlights
Feed-forward	Rummelhart & McClelland [8]	Back-propagation	– Supervised learning: poor reflection of human lexical acquisition – The earliest architecture defined – Able to capture only a highly limited range of phenomena – Inadequate to capture temporal dimension of language
Interactive activation	TRACE [9]	Preset weights	– Multi-level architecture – Interactions among different abstraction levels – Competition and cooperation among candidate hypotheses through inhibitory synapses – Temporal context captured by interconnecting multiple copies of the network
SOM-based	DISLEX [17] SOMBIP [10] DevLex [11]	SOM learning + Hebbian learning	– Unsupervised learning: reflects main mode of human language learning – Self-organization allows for emergence of lexical categories – Interaction among different levels of language – Distributed encoding for word semantics based on contextual features – Capture a wide range of phenomena

This model has received a number of critiques, among them:

- that it is not a valid model of language acquisition, because the direct mapping from phonological forms of verb roots to past tenses is considered in isolation from the rest of the language;
- criticisms about the features chosen for representation (that Wickelphones tend to favour positively the aspects of data that convey most information);
- that the results obtained fall short of being generalizable (due to relatively low performance);
- and that the training and testing procedures were unrealistic, as a result of an excessive zeal in modeling the U-shaped learning curve.

4.2 TRACE: A Model of Speech Processing

TRACE by McClelland and Elman [9] is a neural model of human speech perception, which implements activation of words in a lexicon through a combination of phonological and phonotactical features. It set a hallmark in connectionist treatment of language by introducing the notion of interconnection among different abstraction levels of language. A particularly interesting characteristic of TRACE is found in its ability to perform word segmentation without an explicit marker, based only on phonetic interactions.

The TRACE model was based on the principle of *interactive activation* (IA), where units are related by connections that exercise either an inhibitory or excitatory action on their neighbours. TRACE has three layers of neurons, each one representing a higher level of abstraction in language: first, phonetic features; second, individual phonemes; and third, words. Connections exist within and across layers. Inhibitory synapses model situations where the items represented by the co-activating units can not co-exist (competition), whereas excitatory ones model items that are somehow related (cooperation). In addition, the temporal dimension is captured by having multiples copies of the whole network, among which neurons are also interconnected.

There is one particular novelty about TRACE that challenged the traditional perception of the scientific community regarding how the brain network is organized. It is that activation between layers in TRACE works top-down (words to phonemes) as well as bottom-up (phonemes to words). It is a matter of debate whether layers of higher abstraction feed information back to lower layers. Another particular characteristic of TRACE is that the connection weights are all preset to account for the desired model of language: the network does not learn.

4.3 SOMBIP: A Model of Bilingual Lexicon Acquisition

SOMBIP is an ANN model by Li and Farkas [10] of how a bilingual lexicon (i.e. a lexicon where words of two languages appear mixed) is acquired by bilingual learners. The network architecture consists of two Kohonen SOM maps, one phonological (SOM1) and one semantical (SOM2), interconnected via associative hebbian links. The network was trained to learn a bilingual English-Chinese lexicon of 400 words (184 Chinese, 216 English), extracted from the CHILDES database [18].

In order to allow the network to create associations between translation equivalents in the two languages that occur in the bilingual lexicon, if the phonological representation of an English (or Chinese) word is presented to SOM1 and it has a translation equivalent in the lexicon, not only the semantic representation of the same English (or Chinese) word is presented to SOM2 coupled with the word form, but also the semantic representation of the translation equivalent in Chinese (or English) is presented.

Emergence of grammatical and lexical categories in the form of visible clusters appears in SOMBIP, with the particularity that the network is able as well to effectively separate words from the two languages. Interference effects between words, both intra-language and inter-language, were verified by presenting the network with a phonological representation and observing the response it triggers in the semantic map, and vice versa. Different levels of the learner's proficiency in one of the languages were modelled by building the word meaning representations for one of the languages from a smaller portion of the corpus. Words from the dominant language tended to occupy a larger area of the semantic map than before, which caused lexical confusion in the disadvantaged language.

4.4 DevLex: A Model of Early Lexical Acquisition

DevLex [11], by Li, Farkas & MacWhinney, is a neural network model of the development of the lexicon in young children, based mainly in the SOM architecture and inspired by Miikkulainen's DISLEX model [17]. The authors observe that most previous ANN models of lexical acquisition have been based on the supervised back-propagation algorithm for training, thus misrepresenting the mainly unsupervised nature of lexical acquisition in children, and most have also failed at modelling the incremental nature of vocabulary acquisition. To address this issues, DevLex introduces through a combination of SOM and ART (Adaptive Resonance Theory, [22]) modes of operation.

The DevLex architecture is composed of two GMAPs (Growing Maps), one phonological map for dealing with phonological information of words (P-MAP) and one semantic map (S-MAP) for dealing with word's meanings. A GMAP is an arrangement that combines both the self-organization properties of SOM, and the ability of ART networks to create new nodes that become representatives of a new class of inputs. The learning process is modelled like a gradual transition between the SOM and ART modes of learning. During SOM mode, the network undergoes reorganization as a result of exposure to the input patterns. In ART mode, the network is allowed to create new units when the input pattern (word forms or meanings) is sufficiently different from all the patterns stored in existing nodes. At any time, showing the network a word form causes a response in the S-MAP, which models *language comprehension*; while showing the network a word meaning causes a response in the P-MAP, which models *language production*.

The results observed concerted three types of phenomena: category emergence, lexical confusion and effects of age of acquisition. The target 500-word vocabulary from the CHILDES database is structured in 4 major lexical categories (nouns, verbs, adjectives and closed-class words). By comparing each S-MAP unit against its 5-nearest neighbours, category compaction was observed in nouns (more than 90%), then in verbs, in adjectives (circa 80%), and last closed-class words. Lexical confusion in the network was evaluated by looking inside individual units, in order to observe how many words of the lexicon were cluttered into the same unit, and over the associative links that relate phonology and meaning. Largely in agreement with the way how this phenomenon manifests in children, it was found that lexical confusion is very high during early stages of high reorganization (in SOM mode), and then decreases steeply just to reach a minimum in ART mode. Regarding age of acquisition, it was observed that, after the network starts operating in ART mode, the earlier a word was entered for learnig, the less it took the network to construct an unique representation for it and with a correct association between form and meaning.

5 Comparison to Statistical Approaches

Statistical NLP has typically concerned itself with problems that depart from the interpretation of language acquisition (or lexical acquisition) addressed in this survey. Rather than modelling the identification and learning of words of a lexicon from phonological and/or semantic contextual information, as the ANN models we have reviewed in the previous sections do, the methods employed in statistical NLP extract a series of lexical, morphological, syntactic and semantic features from text documents, in order to apply them to higher-level tasks where the focus is on performance on the task at hand, as opposed to interpretability of the results, or imitation of biological or cognitive processes. Examples of such tasks are text categorization, information extraction, machine translation or word sense disambiguation, among others.

An outstanding difference regarding the way features extracted from words and text are employed in statistical NLP with respect with the connectionist models we have seen here, is that in statistical NLP information flows only from lower-level features (i.e. levels of language) to higher-level problems. Lower levels of language (e.g. morphological) are used to solve the NLP problems in the higher levels (e.g. syntactic or semantic). There is no notion of information flowing forward and backwards across language levels (as in the TRACE model in [9]) or interaction (in a hebbian sense) among different levels.

Although some research literature on the application of ANNs to NLP tasks in general has been published in recent years, the overwhelming majority of instances have employed either statistical or symbolic approaches. This scarcity of ANN-related publications might be due, at least in part, to two of the limitations that are usually associated to ANN models: that of the difficult interpretability of results, and the excessive tuning of the network architecture and learning parameters that is required. Moreover, it is hard to integrate existing background linguistic knowledge for use by an ANN, if so desired. Nevertheless, and as mentioned in the introduction, some connectionist models have been reinterpreted within the framework of probability theory. Among the frequently quoted advantages of this reformulation, we find: the possibility of defining principled model extensions, and the explicit addressing of the model complexity problem. Among the disadvantages: the likely increase of computational effort, and the requirement of data distributional assumptions that might hamper biological plausibility.

With respect to the effect of lexical category emergence of which we have seen occurrences in SOM-based models, and which has an intrinsic interest from the standpoint of modelling cognitive processes, a similar category separation could have been attained by resorting to traditional statistical clustering techniques. Nevertheless, models based on SOM offer an additional value concerning analysis and visualization of the resulting clusters, which is afforded by the reduction-of-dimensionality characteristic of SOM.

6 Conclusions

Throughout this brief survey, we have seen a variety of neural network connectionist architectures that can be used to capture phenomena that arise in language acquisition. Arguments that have been raised for and against ANNs as valid models of language acquisition have been presented. This has lead to the explanation of two particular instances of ANN-based models for lexical acquisition, and has enabled us to prove the point that such full-scale neural models as DevLex (for early lexical acquisition) and SOMBIP (for modelling the acquisition of a bilingual lexicon) can reproduce a variety of phenomena that have a parallel in empirical evidence: lexical confusion, interference between languages, effects of proficiency and learning capacity, etc. In fact, ANNs such as these, as well as other machine learning methods (as in [23]), provide a computational basis for certain biological and psychological explanations of empirically observed phenomena.

References

1. Quartz, S. R., Sejnowski, T. J.: The Neural Basis of Cognitive Development: A Constructivist Manifesto. Behavioural and Brain Sciences **20** (1997) 537–596
2. Chomsky, N.: Aspects of the Theory of Syntax (1965). MIT Press, Cambridge
3. Quinlan, P. T., van der Maas, H. L. J., Jansen, B. R. J., Booij, O., Rendell, M.: Re-Thinking Stages of Cognitive Development: An Appraisal of Connectionist Models of the Balance Scale Task. Cognition, in press (2006)
4. MacKay, D. J. C.: Bayesian Methods for Back-propagation Networks. In Domany, E., van Hemmen, J. L., and Schulten, K. (eds.), Models of Neural Networks III, Ch. 6 (1994). Springer, New York.
5. Neal, R.: Bayesian Learning for Neural Networks, PhD thesis (1994). University of Toronto, Canada
6. Bishop, C.: Neural Networks for Pattern Recognition (1995). Oxford University Press, New York
7. Friston, K.: A Theory of Cortical Responses. Philosophical Transactions of the Royal Society, B, **360** (2005) 815–836
8. Rumelhart, D., McClelland, J.: On the Learning of the Past Tenses of English Verbs. Parallel distributed processing: Explorations in the microstructure of cognition **2** (1986) 216–271
9. McClelland, J. L., Elman, J. L.: Interactive Processes in Speech Perception: The TRACE Model. Parallel distributed processing **2** (1986) 58–121
10. Li, P., Farkas, I.: A Self-Organizing Connectionist Model of Bilingual Processing. Bilingual Sentence Processing (2002) 59–85
11. Li, P., Farkas, I., MacWhinney, B.: Early Lexical Development in a Self-Organizing Neural Network. Neural Networks **17** (2004) 1345–1362
12. Christiansen, M. H., Chater, N.: Connectionist Natural Language Processing: The State of the Art. Cognitive Science **23** (1999) 417–437

13. Pinker, S., Prince, A.: On Language and Connectionism: Analysis of A Parallel Distributed Processing Model of Language Acquisition. Cognition **28** (1988) 73–193
14. Garson, J.: Connectionism. The Stanford Encyclopedia of Philosophy (2002). Edward N. Zalta (ed.)
15. Fodor, J. A., Pylyshyn, Z.: Connectionism and Cognitive Architecture: A Critical Analysis. Cognition **28** (1988) 3–71
16. Pollack, J. B.: Recursive Distributed Representations. Artificial Intelligence **46** (1990) 77–105
17. Miikkulainen, R.: Dyslexic and Category-Specific Impairments in a Self-Organizing Feature Map Model of the Lexicon. Brain and Language **59** (1997) 334–366
18. MacWhinney, B.: The CHILDES Project: Tools for Analyzing Talk (2000). Mahwah, NJ: Lawrence Erlbaum Associates
19. Kohonen, T.: Self-Organizing Maps. Springer Series in Information Sciences **30** (1995)
20. Elman, J. L.: Finding Structure in Time. Cognitive Science **14** (1990) 179–211
21. Anderson, B.: Kohonen Neural Networks and Language. Brain and Language **70** (1999) 86–94
22. Carpenter, G. A., Grossberg, S.: The ART of Adaptive Pattern Recognition by a Self-Organizing Neural Network. Computer **21(3)** (1988) 77–88
23. Yu, C., Ballard, D. H., Aslin, R. N.: The Role of Embodied Intention in Early Lexical Acquisition. Cognitive Science **29** (2005) 961–1005