Connectionist Learning to Read Aloud and Correlation to Human Data

Ivelin Stoianov (stoianov@let.rug.nl)
Faculty of Arts, University of Groningen, POBox 716, 7600 AS The Netherlands

Laurie Stowe (l.a.stowe@let.rug.nl)
Faculty of Arts, University of Groningen, POBox 716, 7600 AS The Netherlands

John Nerbonne (nerbonne@let.rug.nl)
Faculty of Arts, University of Groningen, POBox 716, 7600 AS The Netherlands

Abstract

A research on connectionist mapping from written to spoken forms in natural language is presented. The more plausible for this task Simple Recurrent Network was used instead of a static one. The model was trained on Dutch monosyllabic corpus. The effects of frequency, length and consistency were examined too and they were found matching to reported data in psycholinguistic experiments.

Introduction

Among a number of linguistic problems attracting attention in cognitive science is the word reading task, in particular the mapping from written to spoken forms in natural language. Connectionist models, if successfully trained on this problem and if their performance correlates to human performance in reading, can supply a framework for lexical processing. For example, Seidenberg & McClelland (1989, hence SM89) and Plaut et al (1996) suggest that a *single-route distributed* process performs this transformation, as opposed to the symbolic *dual-route* model which claims that the reader must also have a lexical route that handles exceptions, or irregular words(Coltheart 1980, 1993). The dual-route model has a connectionist implementation too: Zorzi (1998) proposed a Multilayered Perceptron (MLP) with an alternative structure to handle both easier rule-based mappings and the more difficult, exceptional words.

Although these connectionist models are reported to perform well, they employ static lexical encoding, which imposes constraints and might be considered a theoretical drawback. A better account for the variable and sequential nature of words would be sequential processing, which produces single phoneme at a time, as in the Sejnowski & Rosenberg's NETtalk model (1987). Simple Recurrent Networks (SRN) by Elman (1990) fit even better in this lexical representation scheme, with the advantage of gradual left context dependence, as opposed to the window context dependence in the NETtalk model. This capacity is due to the distributed contextual memory in SRNs, gradually evolving in time, while in the NETtalk the temporal information is encoded in a fixed-size window.

SRNs are reported to be successful in other difficult lexical tasks, e.g., learning the phonotactics of monosyllabic Dutch corpus (Stoianov, Nerbonne & Bouma 1998, hence SNB98), which raises hopes for success in the reading task. Also, Plaut (*in press*) employs an extended SRN model for this task. In the current paper we propose further exploitation of the SRN model on this challenging problem.

To accomplish this, we trained a SRN on mapping orthographic-phonetic representations of all 6100 Dutch monosyllables, as found in the CELEX lexical database. This difficult data set contains also rare and foreign words. The lexical encoding that was used in phonotactics learning (SNB98), in which the left context only is presented, is not enough for this task, because only the proper output phoneme should be activated, as opposed to predicting all successors in the phonotactics problem. Therefore, a more specific data presentation was applied. The learning was successful and the network generalized well too. In addition to learning, we examined the SRN's errors for various effects found in previous psycholinguistic experiments, such as word frequency and grapheme-to-phoneme mapping consistency. The sequential data representation allowed us to observe other effects too, such as word length and error positioning. The SRN error profile matched closely to the human performance in reading, which supports the suggestion that SRN models can be used as a basic sequential processing module in a larger cognitive framework, explaining our cognitive linguistic capacity.

Mapping from orthography to phonology & the evolution of connectionism.

The reading process is complex. It involves acquiring visual input; transformation to abstract graphemic representations; further mapping to abstract auditory representations and finally, production of motor commands that cause sounds. If the visual data has semantic meaning, it is accessed too. Simultaneous modeling of all these stages is difficult, so assuming that the input and output steps are done, one can work on the intermediate level. Our work involves only the

mapping from abstract graphemic representations to abstract phonetic representations. Therefore, hereafter by 'reading', we will mean this mapping.

Since the first successful connectionist system that transformed text to phonemes – the NETtalk (Sejnowski & Rosenberg 1987) – a number of other connectionist architectures that model the human reading process have been suggested. Among these, Seidenberg & McClelland (1989) and Plaut, McClelland, Seidenberg & Patterson (1996), have been influential on connectionist NLP research (e.g., Milostan & Cottrell 1998; Harm 1998). A connectionist model that takes the opposite side of the *single route* / *dual-route* controversy was proposed by Zorzi (1998), who suggests that the MLP might benefit from an extra set of connections, from the input to the output layer. This extra set of connections is interpreted as a second "route", similar to the Grapheme-to-Phoneme-Conversion (GPC) rules (Coltheart 1978). The information flow through the standard hidden layer is expected to interpret the more complex mappings, or exceptional words.

All these models are based on the general PDP postulates of using distributed representations and distributed knowledge, principles that stem from our neural system. Connectionist models benefit other useful properties too – generalization for unseen data, noise resistance, etc. – which are considered hard for symbolic systems.

An important divergence between NETtalk and the latter models is the employed lexical representation. In NETtalk, which is a static feedforward MLP, words are represented sequentially to the network, one letter at a time, together with a context of a few letters surrounding the letter to be pronounced. The network is trained to produce the phoneme that corresponds to the current letter and context. In contrast, the models in SM89, Plaut et al. (1996) and Zorzi (1998) explore static lexical representations, where words are presented to the NN at once and the corresponding phonologic representations are produces at once, too.

The SM89 model uses representation based on triples of graphemes and phonemes: "Wickelfeatures". In the input pattern correspondent to a given word, active orthographic Wickelfeatures will be those, which are a sub-part of the input word. The output phonetic encoding is similar. The model used 400 input orthographic units and 460 output phonetic units. This representation raises the necessity of a complex feature encoder and decoder.

The connectionist models proposed in Plaut et al (1996) are a feedforward MLP and an attractor NN (an extension of MLP with a recurrent layer at the output, which aimed at more precise targeted identification). They explore an alternative static data representation, which accounts for the spelling-to-sound regularities of English monosyllabic words. In this representation, there are slots for the onset and coda consonants and the vowels of the nucleus. By observing the existing graphotactic and phonotactic restrictions in the orthographic and phonetic onset, nucleus and coda representations, the authors manually constructed reduced input and output representations with 105 input grapheme and 61 output phoneme units. These units stand for a limited number of orthographic and phonetic onsets, nuclei and codas. After a number of successful experiments on learning a orthography to phonology mapping, their claim is that this connectionist model and lexical representation can account for the basic abilities of skilled readers to pronounce correctly both regular and exceptional items, while still generalizing to novel items. In addition, network error profile with respect to word frequency and grapheme-to-phoneme mapping consistency were tested against human performance in reading. For this purpose, reading latencies were considered to correspond to network error (in MLP) or to the time required for the network to settle to a stable output pattern (in attractor NN model). The experiments and mathematical analysis explain how the networks succeeded in handling quasi-regular domains (both regular and exception words) and producing frequency and consistency interactions exhibited by humans. In spite of this, in natural languages language objects are dynamic, including the words in their orthographic and phonetic representations, which has been dismissed by unjustified application of one of the Hinton' s principles about connectionist models:

> For processing to be fast, the major constituents of an item should be processed in parallel. (Hinton 1990)

Words have constituents (graphemes or phonemes) that are inevitably encountered in a strictly sequential fashion, therefore, words should be processed sequentially. Phonemes span time too, but dealing with words, we can consider the phonemes as static objects and process them in parallel. As far as the graphemes are concerned, one might argue that in reading we perceive visual objects larger than single graphemes, e.g., words. But this is done by skilled readers, possibly by use of some extra mechanisms. Beginners initially read one letter at a time (or group of letters) therefore the models should account for this.

Given this, there is nothing wrong with the NETtalk model, which was criticized in Plaut et al (1996) because of the Hinton's principle. Nevertheless, NETtalk has another problem, based on the fixed limited context. The network would not map correctly two words with different phonetic representations, which differ somewhere beyond the temporally shifting graphemic context scope. A model that is theoretically able to handle such dependencies is Simple Recurrent Networks (Elman 1990), where the output of the network depends on the whole left context, and which we explore in the following section.

Learning to Read Aloud with SRN

The experimental setting in our research is based on a standard SRN, trained to learn sequential association in orthography-to-phonology conversion (Fig.1). The network performance was measured on the training and unseen words. Further, the performance was analyzed for different variables, such as word frequency, length, grapheme-to-phoneme mapping consistency and error positioning, from which we drew some conclusions about the syllabic

structure in Dutch. Preliminary analysis of a damaged network and its correlation to dyslexia is provided too.

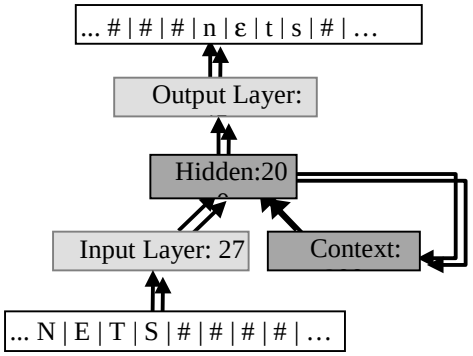

Fig.1 Simple Recurrent Networks and mapping from orthography to phonology.

Method

A Simple Recurrent Network (Elman 1990) was trained to learn a sequential mapping: the orthography-to-phonology conversion of all 6100monosyllabic Dutch words, as extracted from CELEX lexical database. The same corpus was used in our previous studies on graphotactics and phonotactics (SNB98). The data set contains orthographic and phonetic word representations and the frequencies of word occurrence in the Dutch language. This corpus was split into a training set (5100 words) and a test set (1000 words). The orthographic and phonetic word representations have mean length 4.53 (σ=1.08, *min*=3, *max*=9), and 3.92 (σ = 0.94, *min*=2, *max*=8) respectively. The word representations are built of 26 graphemes and 44 phonemes, plus one extra symbol representing space ('#'), used as a filler specifing end-of-word. The graphemes and phonemes were encoded orthogonally, that is, for each grapheme and phoneme, there is one input or output neuron respectively. One might speculate that with regard to the network associative capacity, this representation is equivalent to a distributed, feature-based representation, because we can always add two more static layers that decode and encode such feature-based representations to the orthogonal ones. During training and testing, the orthographic and phonetic word representations were given to the network sequentially, one symbol at a time. In order for the network to be able to correctly reproduce different phonological representations for words with identical beginnings (left context), phonological production is delayed for three steps. In this manner, the network receives partial right context as well. Therefore, the network decision at each moment is based on the full left and partial right context (3 graphemes), simultaneously encoded distributively in the context layer.

Experiments were conducted with different number of hidden neurons, ranging from 100 to 400. Best performance was found with the largest network. In this report, we present results for a network with 200 hidden neurons, resulting in about 55,000 weights. The network architecture is given in Fig.1, where an example mapping from the orthographic to the phonetic representation of the word 'nets' is shown as well. The delay is implemented technically by producing three filling symbols (end-of-word - '#') at the output and feeding the input with the same symbol until the full phonetic representation is generated at the output layer.

Training

The training process is organised in epochs, in the course of which the whole training data set (5100 words) is presented to the network in accordance with word distribution, that is, word frequencies (SNB98). In order to reduce the learning time, the actual word frequencies were shrunk by applying a logarithm function, resulting in about 12,500 training sequences per session. Such an approach has been used by other authors as well (e.g., Plaut et al 1996; Zorzi 1998). Next, for each word, the sequence of graphemes is presented to the input, one by one, followed by three end-of-word symbols. Each time step is completed by copying the hidden layer activations to the context layer, which are used in the next step (Elman, 1990). At the same time, after the network generates its expectations for the phonemes at the output layer, the representation of the true phoneme is used to compute an error for the current time step. This error is used by the BackPropagation Through Time (BPTT) learning algorithm (see for details Haykin 1994; SNB98), which includes a forward move where errors are collected and a backward move, during which global error is back-propagated through time until the beginning of the current training sequence. This process is followed by updating the network weights with values, accumulated during the backward move. The state of the network (i.e., the context memory) is reset after processing one word.

The network was trained on 18 epochs, resulting in approximately 200,000 word presentations. The total number of individual word presentations ranged from 18 to 200, according to the individual word frequencies. The network started with a sharp error drop to about 4%, slowly decreasing down to 1.2% (see Table 1).

Table 1. Dynamics of the SRN error during the training.

Epoch	1	2-4	4-10	10-17	18
Error (%)	4.1%	2%	1.6%	1.4%	1.2%

We expect further error decrease with longer training, although this would need much more training time, because the learning coefficient η decreases by 30% after each epoch, starting from 0.3 and restricted with a bottom limit at 0.001. Learning grapheme to phoneme conversion is quite a difficult task, so we had to apply some other special techniques to improve it. First, instead of the standard Backpropagation algorithm, we used a BPTT learning scheme as described above. Next, standard momentum α=0.5 term was applied. Further, the training process was supervised by an evolutionary algorithm that trained a pool of networks on the same problem and after each training epoch, it eliminated the network with the worst

performance, keeping clones of the networks that performed better. This training method was developed in our previous studies on phonotactics (SNB98) and was found to perform better than the standard single-network training.

Performance

A procedure that examined all training examples followed each training epoch, in which we distinguished phonemic and network errors. *Phonemic* error occurred if during the network processing, the most active neuron didn't correspond to the expected phoneme. *Network* error estimated the percent of all mispronounced phonemes, weighted by the frequency of the word the phonemes belong to. This procedure results in fear estimation of the network performance, accounting for the distribution of the words in natural language. As we mentioned earlier, the final network error estimated during the training was 1.2%. We analyzed the type of the incorrect productions the network has made and found that 75% of them were substitutions between close phonemes, mainly vowels, which might be used as an argument to reduce error.

The generalization capabilities of the network were tested on a test set, which contained the orthographic and phonetic representations of 1000 unseen during training words. For the sake of comparison to Plaut et al (1996), we should note, that the test words should be interpreted as *nonwords* because they have not been used for training. We could use them for testing, because we had their correct phonetic representations. Still, there might be words that are exceptional with regard to the reading, therefore we expected higher error. The performance on this test set was 1.4%, which confirmed that the network learned the Dutch GPC rules for monosyllables. As we predicted, error increase was primary due to the exceptional words.

The overall performance is similar to Zorzi (1998) and worse than SM89, and Plaut et al (1996), which we attribute to the twice larger data set and incomplete training. Obviously, the 18 training epochs resulting in 18 up to 200 exposures for a single word were not enough to achieve perfect performance, especially for the exceptional words. Also, networks with larger hidden layers tend to learn better, however at the cost of longer training time.

Error profile analysis

In addition to overall network accuracy, connectionist systems that model lexical tasks also aim at approximating the correspondent human performance with regard to different variables such as word frequency. This aspect in connectionist modeling is important, because it contributes to verifying whether the suggested models can be used for modeling the correspondent human processes. For this purpose, the model performance is compared with reaction time or error in reading.

The variables we examined were word *frequency*, word *length*, *consistency* of the grapheme-to-phoneme mapping and *error positioning*. Previous reports (Plaut et al 1996; Zorzi 1998) deal mainly with word frequency and consistency, unable to exhibit significant length effects. The sequential nature of SRNs and structure of the training process naturally involve these characteristics, so we were able to test them, as does Plaut (*in press*).

Consistency in orthography to phonology conversion measures how much the pronunciation of a given item is coherent to the pronunciation of orthographically similar items. An interesting issue is how to measure consistency. Plaut et al (1996) used the similarity in spelling of *rhymes* (see below) in order to estimate it. We adopted another definition, suggested by Jared et al (1990), according to which consistency depends on the summed frequency of the word's *friends* and word's *enemies*. "Friends" are words with similar spelling and similar pronunciation, while "enemies" are words with similar spelling, but distinct pronunciation. We categorized consistency into four categories (similarly to Plaut et al 1996). Words with many more friends than enemies are called *regular*, as opposed to words with many more enemies than friends, which are named *exceptions*. There are two intermediate categories – *ambiguous* – with as many friends as enemies – and *semi-regular* – with somewhat more friends than enemies. Error with regard to consistency is given in Table 2. The strong interaction between error and consistency is in parallel to the observed effect of reduced naming latencies and greater accuracy in pronunciation for regular words and increased latencies and lower pronunciation accuracy for irregular ones (Coltheart 1978; Glusko 1979). Still, we see in Table 2 much higher error for exceptional words, which we attribute to the insufficient training.

Table 2. SRN performance against word consistency.

Consistency	Exception	Ambiguous	Semi-Regular	Regular
Error (%)	30 %	5 %	0.8 %	0.1 %
Entropy	1.4	1.2	1.05	1.0

The next important effect we verified was the network performance for different word frequencies (Table 3). This effect was observed in most of the models (Plaut 1996, 1998; Zorzi 1998) and psycholinguistic studies; The SRN also exhibited good frequency effects with up to five times better performance for high-frequency words as compared to the low-frequency items.

Table 3. SRN performance against word frequency.

Frequency	Low	Mid-Low	Mid	Mid-High	High
Error (%)	4.1%	1.5%	1.0%	0.7%	0.8%

In previous studies (Plaut et al 1996 for review) important frequency-consistency interaction was found, where the frequency effect almost disappears for consistent words. To test for this effect, we conducted 2-dimensional analysis of error with regard to frequency and consistency and found the pattern exhibited in human reading studies (Fig.2). Frequency is unimportant for consistent words, somewhat important for ambiguous words and crucial for exceptional words. We should note that the significant error for very

exceptional words doesn't influence much the overall error, due to the very small number of words from that category.

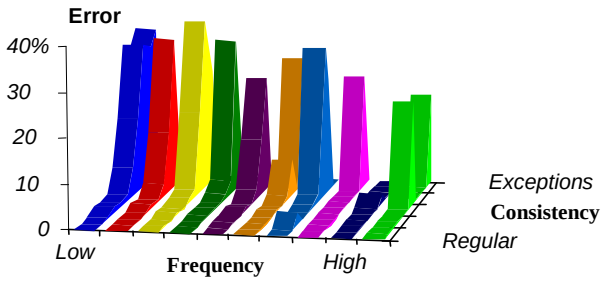

Fig.2 Network error for various degrees of grapheme-to-phoneme consistency as a function of frequency & consistency. The frequency influence on error almost disappears for regular words, where error is very small.

Further, we analyzed the network performance with regard to word length. Connectionist models based on static word representations didn't show apparent interaction between error and length, because static word processing can not be affected by word length. This was the reason Plaut to turn to a sequential connectionist model in his recent paper (*in press*). As we see in Table 4, there is a specific U-shaped dependency between error and word length. For short and long words, error is higher than for words of size 4 to 7. Our explanation for the higher error of the short words is that the network doesn't have enough information to produce the correct pronunciation. On the other hand, the higher error for longer words is easy to explain, as error accumulates during left context processing.

Table 4. SRN performance against word length.

Length	Short	Mid-Short	Mid	Mid-Long	Long
Error (%)	11%	3.1%	2.5%	3.0%	5.6%

And finally, we present an error analysis with regard to the position at which it appears during the process of phoneme producing (Table 5). Given the higher error for longer words one might expect that higher error at the final positions, but it is not so clear why it is higher at the 2nd and 3rd positions. In order to answer to this question, we conducted further finer-grained analyze of the position of the error concentrating on sub-lexical units.

Table 5. SRN performance against error position.

Position	1	2	3	4-5	6-7
Error (%)	1.9%	4.8%	3.7%	1.7%	2.5%

Syllabic structure

Most linguists divide syllables into *onset* (which contains the initial consonants), followed by *nucleus* (intermediate vowels) and ending with *coda* (the final consonants). Also, they look for a more complex internal structure of the syllable, e.g., (1) or (2) (Kessler & Treiman 1997), most theoretical linguists preferring (1).

(1) (onset – rhyme(nucleus – coda))

(2) (body(onset – nucleus) – coda)

Observing the non-uniform error distribution in the words (see above), we were interested how the error was distributed into the above sub-syllabic units. The results showed that the higher error in the middle of the words was due to higher error at the vowel positions (which includes positions 2 and 3 in Table 5 due to variable onset length). The error at the vowel was 8%, while the error in the coda was about 1.9%. There was lower error in the onset as well (1.5% – 1.8%). This means that there is a specific error break at the transition onset-nucleus. This error peak is in parallel to another interesting fact that we observed, namely that the mean entropy in the body was 3.37, σ=0.55, while the mean entropy in the rhyme was 1.90, σ=1.21. This supports statistically the position that the body is less coherent, and the rhyme more.

Therefore, we can conclude that the syllabic structure in Dutch is not plain, but follows the onset-rhyme division (1). The same structure was found in English as well (Kessler & Treiman 1997). With a non-sequential model this would be difficult to measure.

Network damaging and dyslexia

The PDP models are well known for their damage resistance due to their distributed way of representing data and knowledge. A network damage might consist of loss of neurons or memory distortion. We experimented with adding noise to weights and neuron removal. The noise was represented as random numbers, uniformly distributed in the range [-p … p]. There was no apparent effect of 1-2% hidden neurons removal, which was similar to adding slight noise (p=0.10) to 90% of the neurons, while removing 5% of the hidden neurons resulted in 20-25% error, which was similar to adding noise with p=0.30. A milder effect was observed with noise p=0.20, where the error jumped to 6%.

We investigated how damages influenced network performance for various frequency and consistency levels, searching for processes similar to those found in dyslexics. There are two main categories of dyslexia resulting from brain damage (Coltheart 1993; Plaut 1996). In *surface* dyslexia, patients can read non-words but have problems with exceptional words – they regularize them by using Grapheme-to-Phoneme-Conversion (GPC) rules (Colthearth 1993). The other pattern, *phonological* dyslexia is characterized by difficulties in pronunciation of non-words, although familiar words can be read, i.e., patients seem to have lost GPC rules. The pattern of error we observed after network damage looks like phonological dyslexia. The exceptional words get worse, but the affect on regular words is more apparent. Mild damage (noise, p=0.20 or discarding 2-3% of the hidden neurons) affects ambiguous (from 4-5% to 15% error) and regular words (from 0.2-1% to 2-10% error) much more than words with exceptional pronunciation. At the same time, the frequency effect was reduced significantly. More severe damage (removing 5% neurons or noise, p=0.30) resulted in much larger error for regular words and a fading of consistency effect.

In addition, we would note that increasing noise affected almost twice as many test words (i.e., *nonwords*) with exceptional pronunciations as training words from the same category; that is, some test words with exceptional pronunciations perhaps assimilate to the reading of some similar words, but a slight amount of noise easily disrupts this relation. Also, weight damage was more likely to affect the end of the words, which we attribute to corrupting the rules that the network has built in order to encode the context in memory.

Discussion

In this paper we presented our initial experiments on learning grapheme-to-phoneme mapping in Dutch with Simple Recurrent Networks. The main goal in the study was to test the ability of SRNs to learn such a complex mapping employing very simple data encoding – sequential presentation of a single grapheme at a time – as opposed to static connectionist models (Plaut et al 1996; Zorzi 1998) and the more complex sequential mapping scheme in Plaut (*in press*). In order to test how well SRNs approximate human performance in reading, we studied the influence of word frequency and consistency, as well as word length and error-position, which static connectionist models could not observe. Further, trained networks were deliberately damaged with an aim to model dyslexia. SRN performed well on training and unseen test data sets even after very limited number of training epochs. Also, there were significant consistency and frequency effects on error. The error interacted with word length as well. The observed pattern of error positioning suggested a specific non-symmetric syllabic structure for Dutch, which was found in English as well. The reported data on damaging did not show all dyslexic patterns, but we should note that the experiments are still in progress and the data is suggestive.

How does this study contribute to the dialog between single- and dual-route models? We consider SRNs as a single-route model, where a single, although complex mapping produces all outputs in contrast to the "dual-route" network by Zorzi (1998). However, is Zorzi's model dual-route? We claim that it can be simulated by a single-route MLP with restrictions on the weights during training. Zorzi's network structure – with connections that map directly from orthography to phonology (standing for GPC rules) and another set of connections that maps through a standard hidden layer (supposed to maintain a lexicon) – can emerge in training. Still, because it is difficult to analyze the way a uniform network does its job, Zorzi's model contributes to our understanding how neural networks can handle such a difficult problem. Maybe, if we structure SRNs in a similar way, we might achieve even better performance, by minimizing the complexity of the learning task. In addition, this would help to model surface and phonological dyslexia in a more direct way.

References

Christiansen, M. & Nick Chater (1998). Toward a connectionist model of recursion in human linguistic performance. *Cognitive Science* (in press).

Coltheart, M. (1978). Lexical Access in simple reading task. In Underwood (ed.) Strategies of information processing, pp. 151-216. London, Acad. Press.

Coltheart, Max, K. Petterson & J.C. Marshall (1980). *Deep Dyslexia. London*, Boston: Routledge & Kegan Paul.

Coltheart, Max, B. Curtis, P. Atkins & M. Haller (1993). Models of Reading Aloud: Dual-Route and Parallel-Distributed-Processing Approach. In *Psychological Review*, Vol.100, N.4, 589-608.

Elman J.L. (1990). Finding structure in time. *Cognitive Science*, 14, 213-252.

Glushko, R.J. (1979). The organisation and activation of orthographic knowledge in reading aloud. Journal of Experimental Psychology: Human Perc&Perf.,5,674-691.

Harm, M. (1998). Division of Labor in a Computational Model of Visual Word Recognition. Ph.D. thesis at UCS.

Haykin, Simon (1994). *Neural Networks*. Macmillan College Publications.

Hinton, G.E. (1990). Mapping part-whole hierarchies into connectionist networks.*Artificial Intelligence*,46(1),47-76.

Jared, D., McRae, K. & M. S. Seidenberg (1990). The basis of consistency effects in word naming. Journal of Memory and Language, 29, 687-715.

Kessler, Brett & Rebecca Treiman (1997). Syllable Structure and the Distribution of Phonemes in English Syllables. *Journal of Memory and Language*, 37,295-311.

Milostan, J.C. & G.W. Cottrell, (1998). Serial Order in Reading Aloud: Connectionist Models and Neighborhood Structure. In *Advances in Neural Information Processing Systems*, 10, MIT, Cambridge, MA.

Plaut, D.C., J McClelland, M Seidenberg & K Patterson (1996). Understanding Normal and Impaired Word Reading: Computational Principles in Quasi-Regular Domains. *Psychological Review*, 103, pp.56-115.

Plaut, David C. (*in press*). A Connectionist Approach to Word Reading and Acquired Dyslexia: Extension to Sequential Processing. *Cognitive Science*.

Seidenberg, M.S. & J.L. McClelland (1989). A distributed, developmental model of word recognition & naming. *Psychological Review*, 96, 523-568.

Sejnowski, T.J. & C.R. Rosenberg (1987). Parallel networks that learn to pronounce English text. *Complex Systems*, 1, 145-168

Stoianov, Ivelin P., John Nerbonne & Huub Bouma (1998). Modelling the phonotactic structure of natural language words with Simple Recurrent Networks. In Coppen, van Halteren & Teunissen, (eds.) *Computational Linguistics in the Netherlands 1997*, Rodopi, Amsterdam, Netherlands, pp. 77-95.

Zorzi, Marco (1998). Two Routes or One in Reading Aloud? A Connectionist Dual-Process Model. In *Journal of Experimental Psychology: Human Perception and Performance*. Vol. 24, N.4, pp. 1131-1161.