

Short title: DEVELOPMENT OF VISUAL WORD RECOGNITION

A Developmental Perspective on Visual Word Recognition:  
New Evidence and a Self-Organizing Model

Stéphane Dufau<sup>1&2</sup>, Bernard Lété<sup>3</sup>, Claude Touzet<sup>1&2</sup>, Hervé Glotin<sup>2&4</sup>,  
Johannes C. Ziegler<sup>1&2</sup> & Jonathan Grainger<sup>1&2</sup>

<sup>1</sup> Aix-Marseille University

<sup>2</sup> CNRS

<sup>3</sup> University of Lyon

<sup>4</sup> University of Toulon

Corresponding author:

Stéphane Dufau, LPC-CNRS & Université de Provence, Centre Saint Charles, 3 place Victor Hugo, case D, Marseille 13331, France.

E-mail : [stephane.dufau@univ-provence.fr](mailto:stephane.dufau@univ-provence.fr)

Phone: +33.(0)4.88.57.68.92

Fax: +33.(0)4.88.57.69.05

### Abstract

This study investigated the developmental trajectory of two marker effects of visual word recognition, word frequency and orthographic neighborhood effects, in French primary school children from grade 1 to 5. Frequency and neighborhood size were estimated using a realistic developmental database, which also allowed us to control for the effects of age-of-acquisition. A lexical decision task was used because the focus of this study was orthographic development. The results showed that frequency had clear effects that diminished with development, whereas orthographic neighborhood had no significant influence at either grade. A Self-Organizing neural network was trained on the realistic developmental corpus. The model successfully simulated the overall pattern found with children, including the absence of neighborhood size effects. The self-organizing neural network outperformed the classic interactive activation model in which frequency effects are simulated in a static way. These results highlight the potentially important role of unsupervised learning for the development of orthographic word forms.

Key words: Reading acquisition, Word frequency, Orthographic neighborhood, Self-Organizing Map.

The aim of the present study was to provide a behavioral and computational investigation of the development of orthographic representations during reading acquisition. As marker effects of orthographic development, we investigated word frequency and orthographic neighborhood size. The word frequency manipulation was chosen as a standard measure of lexical influences during reading, whereas orthographic neighborhood size was chosen to provide a more direct measure of orthographic influences as opposed to more general lexical influences (including phonology and semantics, for example). These marker effects are discussed next.

### Word Frequency

The word frequency effect is perhaps the most stable phenomenon in psycholinguistics. It reflects the fact that words that occur more frequently in a given language are processed more rapidly and more accurately than words that occur less frequently (Balota & Chumbley, 1984; Connine, Mullennix, Shernoff, & Yelen, 1990; Forster & Chambers, 1973; Rubenstein, Garfield, & Millikan, 1970). This effect has consistently been found in a variety of tasks (lexical decision, naming, perceptual identification, semantic categorization, for review see Monsell, 1991). The word frequency effect has become one of the hallmark effects of word recognition that any model must account for.

In the earliest models of word recognition, frequency effects were implemented as either frequency-dependent thresholds, frequency-dependent resting level activations, or frequency-dependent search priorities (Forster, 1976; McClelland & Rumelhart, 1981; Morton, 1969; Paap, Newsome, McDonald & Schvaneveldt, 1982). The frequency effect does not naturally “fall out” of these models - it is put there to begin with (see Norris, 2006, for discussion of this point). In other words, frequency effects in these models are static. Clearly, this is a major limitation of these models because, in real life, frequency effects likely reflect a dynamic reactivity following actual encounters with written and/or spoken words. Therefore, accounting for frequency effects with static mechanisms is, at best, an approximation. In this respect, more recent connectionist models (e.g., Harm & Seidenberg, 1999, 2004; Plaut, McClelland, Seidenberg & Patterson, 1996; Seidenberg & McClelland, 1989; Zorzi, Houghton & Butterworth, 1998a) present a major improvement over traditional word recognition models because frequency effects are not hard-wired but result from simulated learning. These models are confronted with a training corpus selected to reflect as accurately as possible the frequency of occurrence of words in written language. Thus, high-frequency words are “seen” more often by the network than low-frequency words, and it is the frequency of presentation of a given word to the network that determines the connection strengths linking the different units that represent that word in the network.

There are however several limitations with current connectionist modeling, mostly tied to the widespread use of back-propagation (e.g., Harm & Seidenberg, 1999; 2004; Plaut et al., 1996; Seidenberg & McClelland, 1989; Zevin & Seidenberg, 2002). First, such models necessitate far greater amounts of training compared with the estimated exposure of children to text during reading acquisition (Hutzler, Ziegler, Perry, Wimmer, & Zorzi, 2004). For example, Saragi, Nation, and Meister (1978) found that words presented fewer than six times were only learned by half of their subjects, while performance jumped to 93% after six presentations or more (see Nagy, Herman, & Anderson, 1985; and Rott, 1999, for similar estimates). Second, these models are subject to “catastrophic forgetting” (McCloskey & Cohen, 1989). That is the network “forgets” previously learned items if these items are not continuously interleaved in the presentation of new items. This arises from the distributed nature of representations in the hidden-unit layer of such models. Third, these models use a “supervised” learning algorithm, while it is clear that the learning of orthographic

representations, as opposed to the explicit learning of grapheme-phoneme correspondences, must at least partially occur in a self-organizing and unsupervised fashion (Share, 1995). Finally, current connectionist models are trained using lexical databases selected to be representative of adult reading habits. Therefore, these databases do not fully reflect the written corpus that children are actually exposed to during their primary school years.

The ideal way to test connectionist learning models would be to use developmental data (Harm & Seidenberg, 1999; Hutzler et al., 2004; Grainger & Jacobs, 1998; Jacobs & Grainger, 1994; Zorzi, Houghton & Butterworth, 1998b). Given our focus on the development of orthographic representations, we were particularly interested in examining effects of frequency and neighborhood in a task that emphasizes orthographic processing, such as lexical decision, rather than reading aloud tasks. However, most previous studies have investigated the development of frequency effects in reading aloud (e.g., Ducrot, Lété, Sprenger-Charolles, Pynte and Billard, 2003; Frith, Landerl & Wimmer, 1998; Sprenger-Charolles, Siegel and Bonnet; 1998). Only a few studies have actually investigated word frequency effects in children using the lexical decision task. For example, Burani, Marcolini and Stella (2002) found that frequency effects were present at each grade level and their size (about 2%) did not vary across grade. However, the Burani et al. study only investigated frequency effects in children aged 8 yrs to 10 yrs (i.e., grades 3 to 5), while we were interested in the development of the frequency effect from the very beginning of learning to read (grade 1).

### Neighborhood effects

The second marker effect of interest was the orthographic neighborhood effect. Orthographic neighbors are words that share all but one letter while respecting letter position (Coltheart, Davelaar, Jonasson & Besner, 1977). The neighborhood size of a word is the number of orthographic neighbors of that word (e.g., “word” has 6 neighbors: cord, ford, lord, ward, work, worm). This is an important variable in modeling because it has effects both at the lexical and sublexical level. At the sublexical level, words or nonwords with many neighbors are processed more quickly because they typically have more frequent sublexical units and orthography-phonology correspondences (e.g., Ziegler & Perry, 1998) or they benefit from stronger lexical feedback (Andrews, 1997). At the lexical level, however, words with many neighbors might be processed more slowly because they suffer from lexical competition and lateral inhibition (Grainger, 1990; Grainger, O’Regan, Jacobs, & Segui, 1989). Indeed, both types of effects have been reported previously. In tasks that emphasize sublexical processing, such as reading aloud, neighborhood effects have been found to be facilitatory (e.g., Andrews, 1989, 1992). In tasks that emphasize lexical processing, such as lexical decision or perceptual identification, some studies indeed reported inhibitory neighborhood effects (e.g., Carreiras, Perea, & Grainger, 1997; Grainger & Segui, 1990; Holcomb, Grainger, & O’Rourke, 2002; Segui & Grainger, 1990), whereas others still reported facilitatory effects (e.g., Andrews, 1989; Ziegler & Perry, 1998). Whether facilitation or inhibition is obtained probably depends on a number of factors, such as the balance between lexical and sublexical processing, task, language, and list composition (for reviews, see Andrews, 1997; Grainger & Jacobs, 1996).

A few studies have investigated neighborhood effects during reading acquisition. One particularly relevant study (Laxon, Coltheart, & Keating, 1988) investigated the neighborhood effect in naming and lexical decision in 2nd and 3rd grade children. They showed that neighborhood size had a facilitatory effect both in naming and lexical decision, with improved performance to words with many neighbors compared with words with few neighbors. Similar results were obtained by Laxon, Gallagher and Masterson (2002), who studied children from 5 to 7 years old in a naming task. Finally, Treiman, Goswami and Bruck (1990)

also showed that nonwords with many rhyme neighbors were pronounced more accurately than nonwords with fewer rhyme neighbors by children in grade 1 and 3. Together then, it appears that developmental studies so far have reported facilitatory neighborhood effects. However, most of these studies used naming tasks and limited the comparison to two age groups.

### Goals of the present study

The aim of the present study was to provide a behavioral and computational investigation of the development of word frequency and orthographic neighborhood size effects during reading acquisition. As outlined above, we will focus on the kind of implicit orthographic learning that is likely to happen in an unsupervised and self-organizing fashion. To specifically investigate orthographic development rather than reading aloud, we employed the lexical decision task. It is well-established that the lexical decision task is much more sensitive to orthographic variables than the reading aloud task (Balota, Cortese, Sergent-Marshall, Spieler, & Yap, 2004; Grainger & Jacobs, 1996). Use of the lexical decision task is further motivated by the fact that the vast majority of typical adult reading activity concerns silent reading, not reading aloud.

From a purely methodological perspective, the present study goes beyond previous studies in a number of ways. First, while previous studies typically compared no more than two age groups, the present study collected data from grade 1 to grade 5. Second, frequency and neighborhood estimates in previous developmental studies were mostly based on adult word counts (for a notable exception, see Ducrot et al., 2003). Instead, in the present study, word frequency and neighborhood size measures were estimated for each grade using a developmental database based on textbooks that are currently used in French primary education for teaching reading (Manulex: Lété, Sprenger-Charolles, & Colé, 2004). Third, the use of this developmental database also allowed us to control for the effects of Age-of-Acquisition (AoA).

Indeed, it has been argued that empirical observations of word frequency effects do not actually reflect an influence of frequency of exposure to words (as will be argued here), but rather the influence of the age at which these words were acquired (e.g., Morrison & Ellis, 1995, see Johnston & Barry, 2006, for review). Nevertheless, a number of studies have shown robust effects of word frequency when AoA is matched across frequency classes (e.g., Bonin, Chalard, Méot, & Fayol, 2001; Brysbaert, 1996). An important aspect of Brysbaert's study is that it used an arguably more realistic measure of AoA, based on teachers' ratings of whether or not children in grade 1 were expected to know a word or not (the majority of studies use adult ratings of AoA). On the other side of the coin, Zevin and Seidenberg (2002) criticized prior research reporting evidence for effects of AoA as having confounded AoA with cumulative frequency. That is, the correct measure of word frequency would not be a single static measure as provided by counts of occurrences in corpora of reading material mostly seen by adults, but rather a cumulative measure of the number of times a word is likely to have been encountered from childhood up to the time of testing in a laboratory experiment. Of course, AoA could still have an influence over and above the effects of cumulative frequency (e.g., Ghyselinck, Lewis, & Brysbaert, 2004), but that is not the object of the present study. Instead, in the present study, AoA was strictly controlled, in that all the words that were tested were already present in the reading textbooks for grade 1.

Finally, from a theoretical perspective, the present study provides one of the first evaluations of unsupervised self-organizing neural networks in the modeling of implicit learning of orthographic representations. Furthermore, while previous connectionist models were typically trained on an adult database, the present network was trained on the same realistic training corpus that was used to establish the frequency norms. Thus, the model was

trained on a corpus that contains the words that children actually encounter during primary school. The network was then confronted with the same words that were used for the experiment. We will first present the experiment and results and then the implementation of the model and the simulations.

## EXPERIMENT

### Method

#### Participants.

One hundred and forty children were pre-tested: 25 were 1st graders (G1), 24 were 2nd graders (G2), 25 were 3rd graders (G3), 41 were 4th graders (G4), and 25 were 5th graders (G5). Participants retained for data analysis were selected on the basis of their reading level which was assessed with a standardized reading test (Alouette: Lefavrais, 1965). Only children at the expected reading level for their grade were retained. Following this criterion, 20 children in each grade were retained except in 4th grade where only 10 children had the appropriate reading level. Mean reading age was 6 years 11 months in G1, 7 years 6 months in G2, 8 years 6 months in G3, 9 years 9 months in G4, and 10 years 6 months in G5.

#### Materials.

A set of fifty-six words of 4 and 5 letters in length were selected from Manulex (Lété et al., 2004). Manulex is a computerized lexical database that provides frequency counts of non-lemmatized and lemmatized words compiled from the 1.9 million words found in the main French primary school reading text books. Manulex provides frequency counts for grade 1, grade 2, grades 3-5 collapsed (because frequency counts vary little across these grades), and all grades combined (G1-G5). Words were selected to fill the four conditions created by crossing word frequency (high vs. low, hereafter noted HF and LF, respectively) with neighborhood size (high vs. low, hereafter noted HN and LN, respectively). There were 14 words in each category. HF targets averaged 517 occurrences per million words (G1-G5 level values), and LF targets averaged 20 occurrences per million words. HN targets averaged 7 orthographic neighbors and LN targets averaged 1 orthographic neighbor. Neighborhood sizes were determined using the standard N metric (Coltheart et al., 1977) applied to the grade 1-5 corpus (Manulex), with the additional constraint that a given word's classification (HN vs. LN) did not change as a function of estimated vocabulary knowledge at each grade. A set of 28 nonwords (pseudowords) were constructed that formed pronounceable, orthographically legal letter strings, and a set of 28 unpronounceable nonwords formed of random combinations of consonants. Each participant saw the entire set of 112 stimuli.

#### Procedure and apparatus.

Children were seated at a fixed distance of 60 cm in front of a 17" color monitor connected to a Pentium III laptop computer running DMDX software (Forster & Forster, 2003). The stimuli were displayed in lowercase in 24-point Courier font with a 640 X 480 resolution. Children were tested individually in a single 25-minute session. Each trial consisted of the following sequence of events. The child was first instructed to look at a fixation point ("+") at the beginning of each trial. After 1000 ms, the fixation point was replaced by a target centered on the screen. The target remained on the screen until the child responded by selecting the word-response (right shift key) or the nonword-response (left shift key on the keyboard). He/she was instructed to respond as quickly as possible, while avoiding errors. The targets were presented in a different random order to each participant. There was one block of 28 practice trials followed by 4 blocks of 28 experimental trials.

### Results

Response times (RTs) and mean percentage of errors to words were calculated across

items (for the by-participant analysis) and across participants (for the by-item analysis) for each grade level and for each experimental condition. Trials with RTs below 450 ms or above two standard deviations of a participant's mean per condition were discarded from the analysis (3.78% of the total trials). A 5 (Grade Level) x 2 (Lexical Frequency) x 2 (Neighborhood Size) ANOVA was conducted with participants (F1) and items (F2) as random factors. Lexical frequency and neighborhood size were treated as a between-item factor in the item analysis. RTs and error percentages are shown in Table 1 (see Figures 2 and 3 for corresponding plots). For a clearer view of the results, simple effects and interactions in the next section are only reported when both F1 and F2 ratios were significant.

**Table 1:** Behavioral Results. Mean Response Times (RT, in Milliseconds) and Error Percentages (Err) on Words from Grade 1 (G1) through Grade 5 (G5). Standard Deviations in Parentheses (HF, high frequency; LF, low frequency; HN; high neighborhood density; LN, low neighborhood density).

	<u>HF-HN</u>		<u>HF-LN</u>		<u>LF-HN</u>		<u>LF-LN</u>		<u>Marginal mean</u>	
	<u>RT</u>	<u>Err</u>	<u>RT</u>	<u>Err</u>	<u>RT</u>	<u>Err</u>	<u>RT</u>	<u>Err</u>	<u>RT</u>	<u>Err</u>
G1	2655 (964)	16 (12)	2389 (717)	20 (16)	3159 (979)	44 (18)	3491 (1150)	53 (16)	2924	33
G2	1325 (459)	8 (6)	1420 (559)	9 (9)	1815 (723)	30 (17)	1976 (716)	35 (19)	1634	21
G3	873 (189)	3 (5)	884 (157)	5 (8)	1091 (262)	21 (11)	1122 (262)	25 (14)	993	14
G4	856 (161)	4 (5)	848 (126)	5 (8)	1026 (306)	19 (10)	1059 (146)	17 (13)	947	11
G5	772 (106)	1 (3)	790 (112)	5 (5)	930 (145)	16 (12)	980 (159)	14 (12)	868	9

*Response times.* There was a main effect of Grade [ $F_1(4, 85) = 54.8, p < .001$ ;  $F_2(4, 208) = 417.9, p < .001$ ] with RTs decreasing as reading skills increased. Children in grade 5 were about three times faster making lexical decisions than children in grade 1. A main effect of Frequency indicated that high frequency words were recognized faster than low frequency words (1357 vs. 1735 ms, respectively) [ $F_1(1, 85) = 156.6, p < .001$ ;  $F_2(1, 52) = 56.5, p < .001$ ]. There was also a significant Grade X Frequency interaction [ $F_1(4, 85) = 17.9, p < .001$ ;  $F_2(4, 208) = 7.8, p < .001$ ] reflecting the fact that the size of the frequency effect diminished with age. The main effect of Neighborhood size was not significant (HN = 1450 ms, LN = 1496 ms) and there were no significant interactions with this variable [ $F_s < 1$ ].

*Errors.* As expected, the ANOVA revealed a main effect of grade level [ $F_1(4, 85) = 25.20, p < .001$ ;  $F_2(4, 208) = 52.80, p < .001$ ], indicating that errors increased with age: there was 34% of errors in G1, 22% in G2, 14% in G3, 12% in G4, and 9% in G5. High-frequency words produced fewer errors than low-frequency words (8% vs. 28%, respectively) [ $F_1(1, 85) = 214.30, p < .001$ ;  $F_2(1, 52) = 32.33, p < .001$ ]. A significant interaction was found between grade level and frequency [ $F_1(4, 85) = 6.10, p < .001$ ;  $F_2(4, 208) = 7.49, p < .001$ ], indicating that the frequency effect decreased as age increased. The main effect of neighborhood size was not significant (HN = 16 %, LN = 19 %) and there were no significant interactions with this variable [ $F_s < 1$ ].

## Discussion

The present experiment examined effects of word frequency and orthographic neighborhood size in primary school children from grade 1 to grade 5. The results of our experiment showed a decrease in the size of the word frequency effect from grade 1 to grade 5, on RTs and errors. On the other hand, orthographic neighborhood size did not significantly influence children's performance.

The word frequency effect observed in first graders in the present experiment was of similar magnitude to the one previously reported by Ducrot et al. (2003). These data therefore suggest that orthographic development, as measured by lexical decision performance, is sensitive to the printed frequency of words from the very beginning of reading development. Although there were no significant effects of orthographic neighborhood, inspection of Table 1 shows that there was a relatively systematic numerical advantage in both RTs and errors for words with many orthographic neighbors relative to words with few orthographic neighbors, which is in line with prior research (e.g., Laxon, Gallagher, & Masterson, 2002; Laxon, Coltheart, Keating, 1988). The fragile nature of the classic neighborhood effect in the lexical decision task is consistent with recent adult studies (Grainger, Muneaux, Farioli & Ziegler, 2005; Mulatti, Reynolds & Besner, 2006; Ziegler & Perry, 1998). In the present study, the neighborhood effect appeared to be limited to the younger readers, although the grade by neighborhood interaction failed to reach significance. The facilitatory nature of the N effect in beginning readers found in prior research, and numerically present in our experiment, could reflect the crucial role of reading by analogy in early reading development (Goswami, 1993; Goswami & Bryant, 1990; 1992).

## SIMULATION

The developmental data described above provide us with ideal material to test our self-organizing neural network model of orthographic learning. It is important to note at the outset that we chose not to simulate phonological recoding or the explicit learning of grapheme-phoneme correspondences (for examples of work on this particular issue, see Hutzler et al., 2004, or Zorzi et al. 1998b). The focus of the present modeling work was on the implicit learning of orthographic representations as an essential, and until now largely ignored, component of the process of learning to read. Thus, our modeling approach consisted of specifying and implementing a set of realistic mechanisms and principles that may underlie the acquisition and representation of orthographic knowledge about words. Our general approach is motivated by the assumption that the spaces surrounding printed words enable accurate association of letter-level information with a unique whole-word orthographic representation, and that this process proceeds largely without explicit supervision. Self-Organizing Maps (SOM; Kohonen, 1982) have the interesting property of being able to represent certain biological cortical activities such as lateral inhibition or redistribution of synaptic resources that occur in an unsupervised manner (Kohonen, 1982; Miikkulainen, 1990). In psycholinguistics, SOMs have been previously applied to simulate phonological development in children (Li, Farkas, & MacWhinney, 2004) or category-specific deficits in semantic representations (Zorzi, Perry, Ziegler & Coltheart, 1999).

Mathematically, SOMs can be seen as a tool for mapping a multidimensional data set onto a much lower dimensional space. An important quality of this mapping process is that frequency information is retained. We therefore expect to observe a word frequency effect during word learning. In addition, similarities across different input vectors are coded in the feature map. The absence of an effect of orthographic neighborhood in the empirical data might therefore be a problem for our model.

SOMs offer several advantages relative to prior attempts to simulate word learning

using backpropagation, over and above the fact that they learn without supervision. First, they require much fewer presentations for successful learning, and as will be shown in the present study, the numbers involved are more in line with the estimated exposure of children to print during reading acquisition. Second, they do not suffer from catastrophic forgetting. The localist nature of higher-level representations in such models protects them against interference from subsequent learning of different items.

## Method

### Learning base.

Three sets of four- and five-letter words were extracted from Manulex (Lété et al., 2004) to serve as the training corpus. The grade 1 set was selected to represent exposure to print in first grade primary education in France, and consisted of 54 014 occurrences of 1198 different words appearing in grade 1 reading text books. The grade 2 set reflected exposure in grade 2 and was composed of 54 341 occurrences with 1511 different words. The grade 3+ training set reflected exposure in grades 3-5 and had 64 007 occurrences with 2404 different words at each grade level. The same training set was used in grades 3-5 because Manulex only provides frequency counts for these three grade levels grouped together.

### Input coding.

SOMs can handle a large variety of different kinds of inputs (Miikkulainen, 2005), and one important quality is their ability to handle “natural” or realistic inputs. The starting point of the present model is an orthographic coding scheme that reflects the kind of flexible, relative-position coding of letters in words that has been highlighted in recent behavioral experiments (e.g., Grainger, Granier, Farioli, Van Assche, & van Heuven, 2006; Perea & Lupker, 2004; Schoonbaert & Grainger, 2004; Van Assche & Grainger, 2006; see Grainger, 2008, for review). Open bigram coding (Grainger & Van Heuven, 2003; Grainger & Whitney, 2004) represents one possible representation of such relative-position coding (see Dehaene, Cohen, Sigman, & Vinckier, 2005, for an account of how open-bigrams can be derived from visual input). Open-bigrams are formed of adjacent and non-adjacent pairs of letters in a given order (e.g., the bigram “ab” implies that the letter “a” is before the letter “b” in the input string). This coding scheme provides accurate order information in the absence of precise, length-dependent, position information. For example, the input of the word TABLE will be represented by the following open bigrams: (TA, TB, TL, TE, AB, AL, AE, BL, BE, LE). In the specific implementation of open-bigram coding used in the present study, two additional steps were taken. First, each bigram activation value was modulated by the number of letters separating the two letters of the bigram in the stimulus word. The factor was equal to 1 for TA, AB, BL, LE (adjacent letters), equal to .6 for TB, AL, BE (letters separated by one letter), equal to .1 for TL and AE (letters separated by 2 letters), and equal to 0 for TE (letters separated by more than two letters). This corresponds to the Overlap Open Bigram model described and tested by Grainger et al. (2006). Second, bigram activation was modified as a function of letter visibility using the empirically determined values provided by Stevens and Grainger (2003). The bigram visibility value was the mean of the visibility values of each constituent letter. For the example stimulus TABLE the letter visibilities were T (.78), A (.70), B (.72), L (.67), E (.65), and so bigram TA had a visibility of .74 (footnote 1). The letters used were the 41 letters of lowercase accented French (26 letters of the Roman alphabet plus accented letters). This produced 1681 combinations of paired letters arranged in a vector. This vector [AA, AB, AC, ..., ZX, ZY, ZZ] was associated with the calculated values of the active bigrams of a given input stimulus and filled with zeros otherwise. Such a vector will be referred to from now on as “bigram-word”. This kind of input defines a huge

multidimensional input space, but several SOM applications involve input spaces of this order of magnitude (e.g., WebSOM: Lagus, Kaski & Kohonen, 2004).

Architecture.

The topology of the two dimensional map (word layer) consisted of 100 x 100 units arranged in a grid. Each unit in this layer was linked to its four map neighbors: North, East, South, West. Each map unit was connected to all the input layer units described above. A part of the network is displayed in Figure 1. The dimension of the inter-layer connection matrix was  $[10^4, 1681]$ , which allows the possibility of single coding for ten thousand different word representations, referred to from now on as “lexical words”. Before learning, each element of the inter-layer connection matrix (the weight matrix) was assigned a real value randomly chosen between 0 and 1.

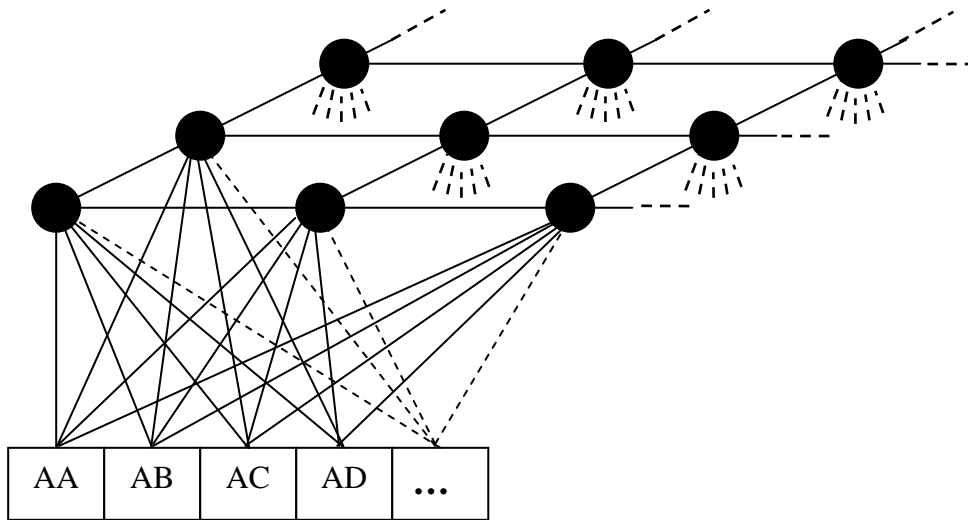


Figure 1: Illustration of Full Connectivity between the Output Map (grid of black points) and the Input Bigram Vector.

Training phase.

SOMs are data-driven models in the sense that the learning algorithm is unsupervised. Therefore our SOM has the ability to learn incrementally from individual bigram-word presentation (i.e., sets of bigrams corresponding to real words) without feedback. Each bigram word presentation leads to connection weight modifications of the maximally responding lexical word unit and its neighbors (see Figure 1). The map used was two dimensional. At the beginning of the learning phase, each map unit contained a random template (1681 real values within the range  $[0; 1]$ , i.e. a weight vector) against which an input (the bigram-word vector described in the input coding section) was matched. When bigram words were presented to the input layer, the algorithm computed in parallel the difference between templates and input for all the map units. The map unit that verified  $\min|W-X|$ , where  $W$  is a map unit weight vector and  $X$  the input values, was selected as the winning unit. Its weights were updated according to

$$W_{t+1} = W_t + \alpha.(W_t - X_t), \quad (1)$$

where  $\alpha \in \mathfrak{R}_{[0;1]}$  and  $t$  the iteration step. The four neighbor unit weights were updated according to

$$W_{N,t+1} = W_{N,t} + \beta.(W_{N,t} - X_t), \quad (2)$$

where  $W_N$  represents the weights of a winning unit neighbor and  $\beta \in \mathfrak{R}_{[0;1]}$ ,  $\beta < \alpha$ . The learning rates  $\alpha$  and  $\beta$  were 0.95 and 0.5 at the beginning of the learning phase, i.e. grade 1 and grade 2 training sets. These values were changed to 0.8 and 0.3 respectively for the grade 3 to 5 training set, which are typical parameter settings (for a more detailed presentation of the SOM algorithm and standard parameter settings see Kohonen, 1995). It is important to note that the construction of the map was incremental. In the first few hundreds of bigram word presentations, only central units responded to the inputs. Due to the properties of the neighborhood weight updating, as learning progressed, responding units spread on the map so that more and more bigram words were processed as unique input.

To ensure a reproducible network response, each training set was presented three times except for the grade 3+ set, which was presented nine times (because this training set comprised grades 3-5). By the end of training up to grade 5, a network had processed about 900 000 bigram word occurrences. Bigram words in each training set were presented in random order, and one hundred and twenty networks were trained according to the procedure described above (24 networks for each grade, which corresponds roughly to 1 network per child). Training was incremental such that networks for grades 2 and above were extensions of the previously trained networks for the lower grades (footnote 2).

#### Test phase and model readout.

The 120 networks were used to simulate the performance of 24 children per grade. Model tests were carried out at the end of the 3 epochs of training for the grade 1 and grade 2 training sets and at the end of the third, sixth and ninth epoch of the grade 3+ training set (grade 3, grade 4 and grade 5 simulations, respectively). The unique maps built at these particular epochs form a representation of the orthographic lexicon of a child at the end of each grade. In order to generate response read-out from the SOM in the form of RTs and percent errors, the bigram-word weights of a given trained network were introduced in a two-layered Interactive-Activation Model (IAM, McClelland & Rumelhart, 1981) with open-bigrams and word representations. The main differences between the standard IAM and the one implemented here are the following: (i) the use of bigrams instead of letters, (ii) the use of SOM weights for setting connection strengths between bigrams and words, and (iii) the absence of resting level activations of the word units at the beginning of a trial. All other aspects of IAM described in McClelland and Rumelhart (1981), including parameter values, were not modified. Simulated reaction times and proportion of errors were then obtained for the 56 words used in the behavioral experiment with the SOM weights corresponding to training after grades 1-5. RTs were measured as number of cycles to reach a criterion activation level (average asymptote minus 10% of this value), and errors were coded as trials on which this criterion level was not reached (i.e., asymptotic activation was lower than the criterion level; footnote 3).

#### Simulation Results

The simulation results are presented together with the human data in Figure 2 (% errors) and Figure 3 (RTs) for both the frequency and the neighborhood effect. Because the error rates in the early grades are quite high (around 50% for low-frequency words), the accuracy data are probably more meaningful than the RT data.

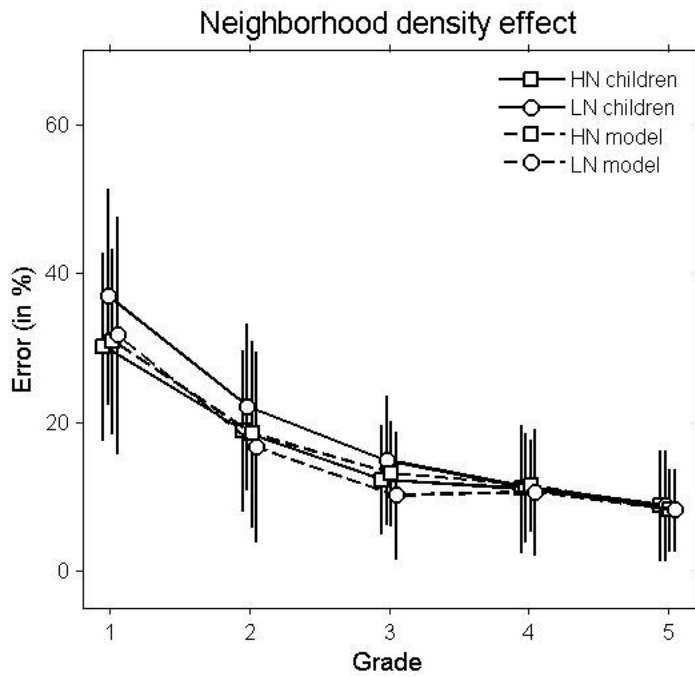
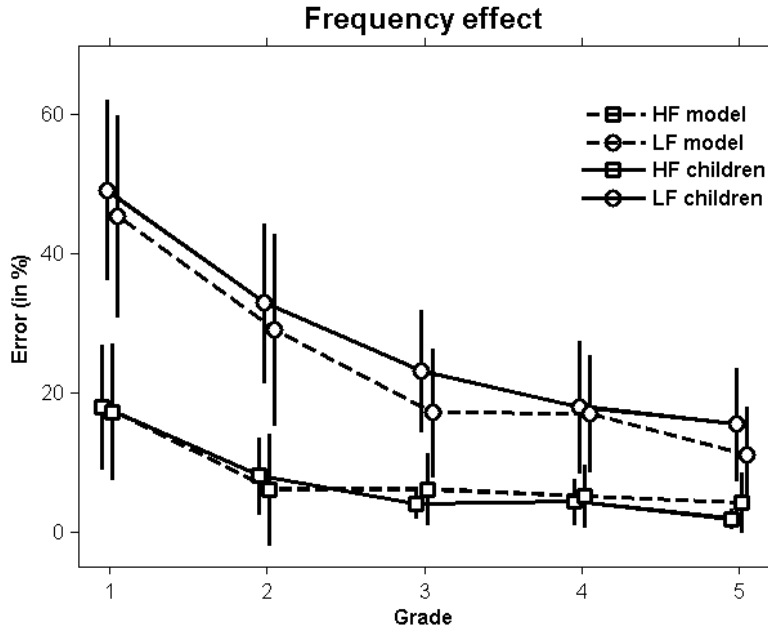


Figure 2: Behavioral and Simulated (SOM&IAM) Effects of Word Frequency (HF – high frequency; LF – low frequency) and Orthographic Neighborhood (number of orthographic neighbors: HN – high; LN – low) on Percent Errors (error bars represent standard errors).

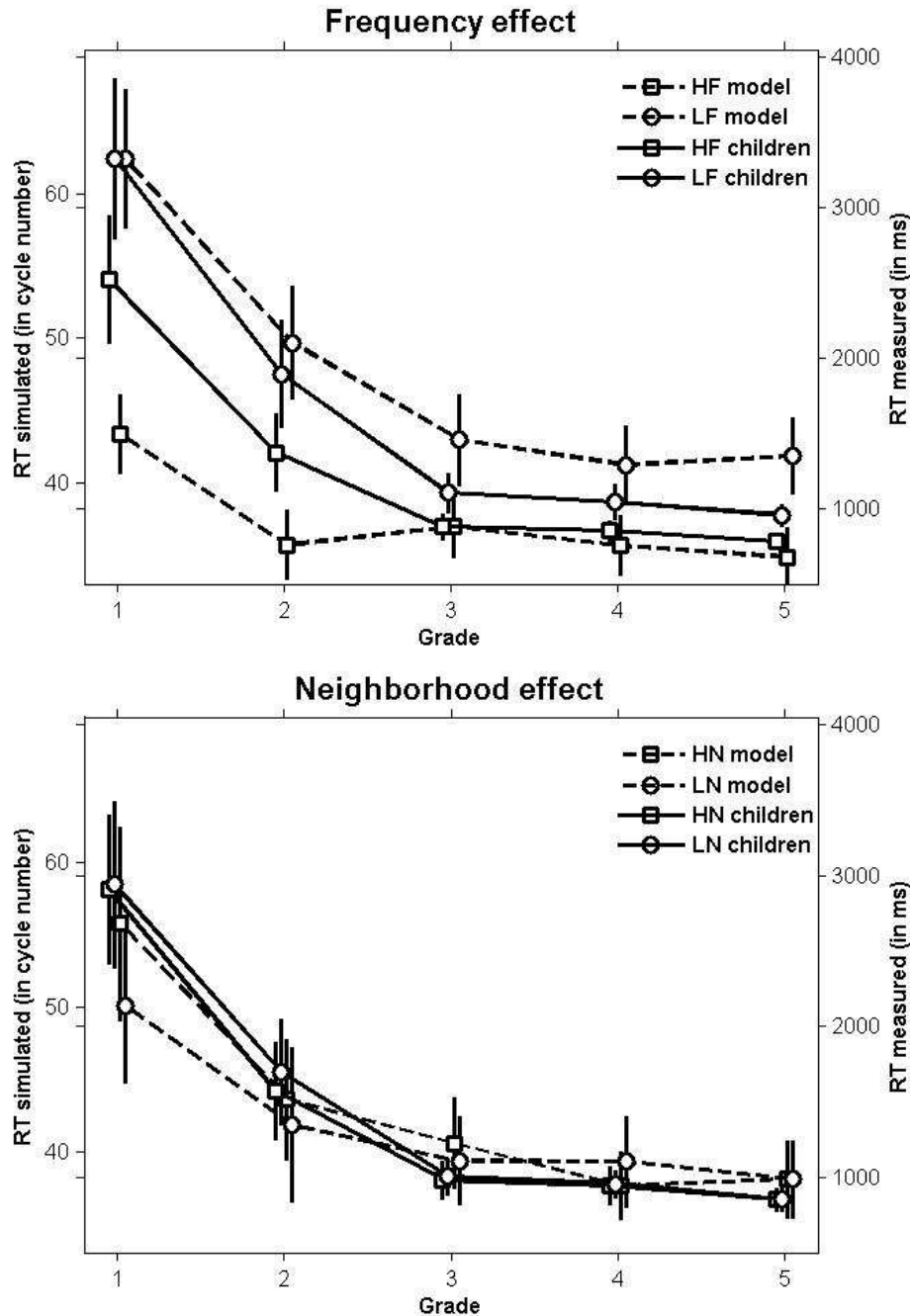


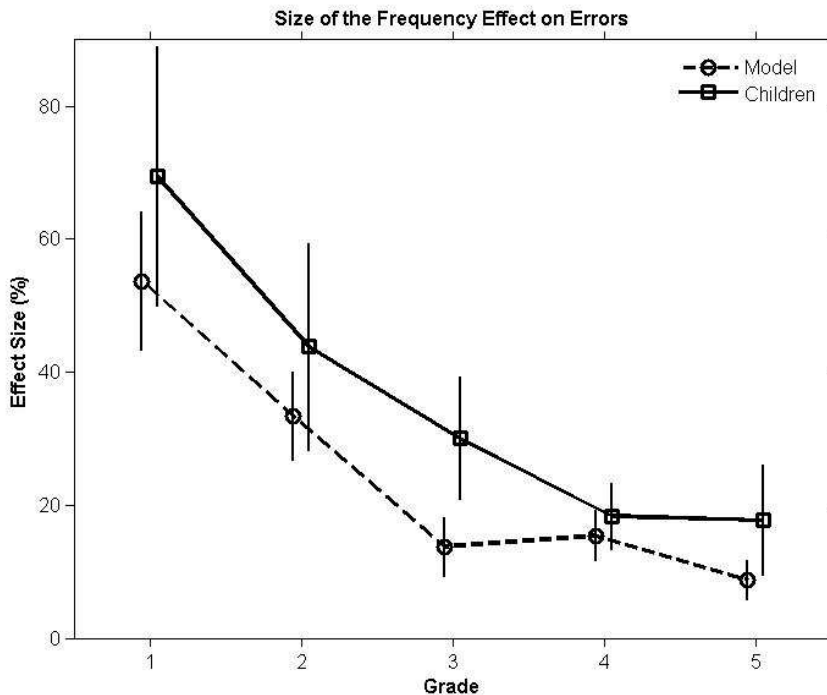
Figure 3: Behavioral and Simulated (SOM&IAM) Effects of Word Frequency (HF – high frequency; LF – low frequency) and Orthographic Neighborhood (number of orthographic neighbors: HN – high; LN – low) on RTs (error bars represent standard errors).

The simulation results were analyzed in exactly the same way as the human data, that is, we performed an ANOVA with Grade Level, Frequency, and Neighborhood Size as factors. The ANOVA was performed on the item means (F2). This analysis revealed a main effect of grade on reaction time (RT) and percent errors (Err), [F2 (4,208) = 18.61,  $p < .0001$ ; F2 (4,208) = 29.12,  $p < .0001$ , respectively], indicating that speed and accuracy both increased with reading ability. A main effect of word frequency was found in both RTs and percent errors, [F2 (1,52) = 16.38,  $p < 0.001$ ; F2 (1,52) = 18.51,  $p < 0.001$  respectively]. High frequency words were processed faster and more accurately than low frequency words. These effects were qualified by a significant interaction between grade and frequency in both RTs and percent errors, [F2 (4,208) = 6.03,  $p < 0.001$ ; F2 (4,208) = 6.60,  $p < 0.001$ ], indicating

that the simulated frequency effect decreased with grade. Again, as in the human data, neighborhood density showed no significant effect, and there were no significant interactions with this factor [all  $F_2 < 1$ ].

### Frequency effects in SOM and children

The SOM successfully simulated the developmental pattern of the word frequency effect observed in children. In order to provide a direct comparison of performance in the model and in children, the frequency effect was transformed into a percentage gain in performance in HF words compared with LF words. These percentage values, shown in Figure 4, were entered into an ANOVA with Grade and Type of Data (Children vs. SOM) as factors. An ANOVA was conducted with participants and simulations ( $F_1$ ) as random factor. An analysis of the error data revealed main effects of Grade and Type of Data [ $F_1(4,200)=42.67$ ,  $p < .0001$ ;  $F_1(1,200)=15.02$ ,  $p < .001$ ], and no interaction [ $F_1 < 1$ ]. An analysis of the reaction time data revealed a main effect of Grade [ $F_1(4,200)=14.02$ ,  $p < .0001$ ], no effect of Type of Data [ $F_1 < 1$ ], and a trend to an interaction [ $F_1(4,200)=2.17$ ,  $p < .1$ ]. This pattern of results indicates that the size of the frequency effect decreased with grade in both the children and the model in a similar manner, and that the frequency effect on RTs was smaller in the model than the effect found in children.



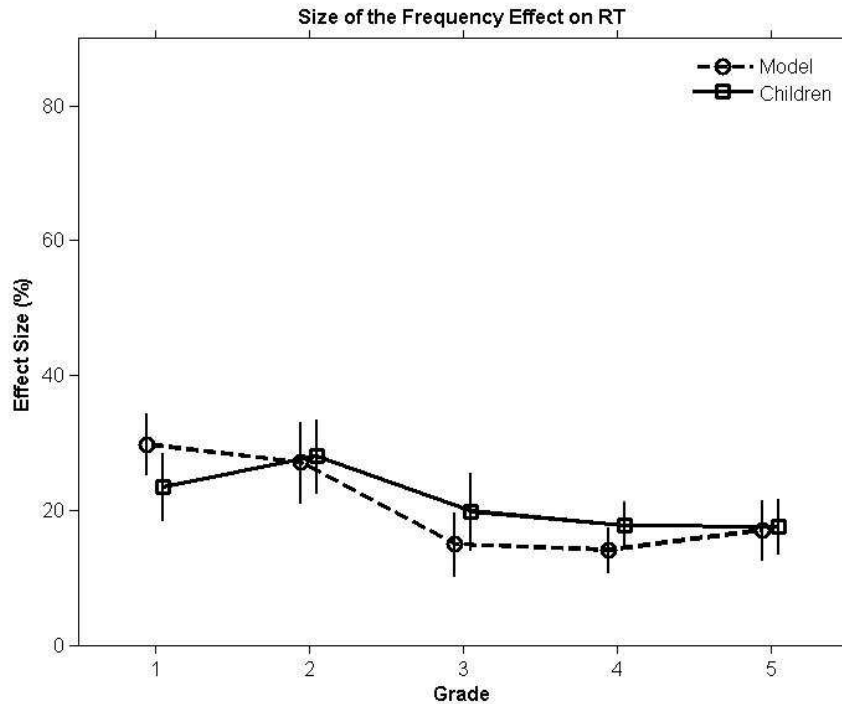


Figure 4: Behavioral and Simulated Effect Size of Word Frequency on Errors and RTs (error bars represent standard errors).

### SOM's contribution

Our modeling approach consisted of building lexical representations via a series of learning stages using a self-organizing map algorithm (SOM). This learning model was combined with the interactive activation model (IAM) in order to generate read-out for lexical decision RTs and errors. The question to be examined now is the extent to which the fit with the behavioral data is driven by the weight adjustment procedure of SOM, or by the architecture and read-out mechanisms of the IAM. In order to dissociate SOM's contribution from the IAM, we compared the performance of the SOM&IAM model with two versions of IAM: one with the standard slot-based letter coding scheme (McClelland & Rumelhart, 1981), and the other with our open-bigram coding scheme for input. In these two models, connection weights between input and word-level representations did not vary as a function of word frequency and were set using the standard parameter values of the IAM. Frequency was coded in terms of variations of resting level activations of word representations, calculated using the cumulative word frequencies across grades as provided by Manulex. Performance of these three models (slot-code IAM, open-bigram IAM, SOM&IAM) was compared by computing correlations on means per item per grade (N=280) for model and children. It can be seen in Table 2 that the open-bigram code version of IAM performed better than the classic slot-code version, hence providing additional support for this type of flexible letter position coding scheme. Most important, the weight adjustment procedure of the SOM produced stronger correlations than the frequency-adjusted resting-level activation model, with everything else being equal between these two models.

Table 2. Correlation between Behavioral and Simulated Results (IAM Slot Code; IAM Open Bigram Code; SOM&IAM Open Bigram Code) for Reaction Time Measures (RT) and

Error Percentages (Err). *p*-levels in parentheses.

IAM Slot Code	IAM Open Bigram Code	SOM&IAM Open Bigram Code	
RT	RT	RT	Err
0.13 (0.02)	0.23 (<0.001)	0.40 (<0.001)	0.47 (<0.001)

The fact that SOM outperformed the standard IAM, with resting-level activations varying as a function of cumulative frequency, is evidence that the pattern of effects found in children is not simply a direct reflection of how cumulative frequency varies across grade level. Indeed, if we plot the actual cumulative frequency values of the low and high frequency words tested in the present study as a function of grade level then we see a much faster rise in cumulative frequency for high frequency words compared with low frequency words. Therefore a simple measure of cumulative frequency would incorrectly predict an increase in the size of the word frequency effect across grade level.

## GENERAL DISCUSSION

The present study provides an empirical and theoretical analysis of the development of the word frequency effect during reading acquisition. Children in grades 1 through 5 were shown high-frequency and low-frequency words in a lexical decision task. Lexical decision performance was found to be significantly more accurate for high-frequency than for low-frequency words, and the size of the word frequency effect decreased significantly with age. Our results show that word frequency has a strong impact on visual word recognition right from the very first phases of reading acquisition. The behavioral experiment also examined effects of orthographic neighborhood density but this variable did not significantly affect lexical decision performance in the present study.

A simulation study with a self-organizing map (SOM - Kohonen, 1982) successfully simulated the main pattern of effects found in the behavioral experiment. The model comprises an orthographic input layer that codes the identity and relative-positions of letters in the stimulus input (open-bigram coding: Grainger & van Heuven, 2003; Grainger & Whitney, 2004), and a “lexical” layer that assigns a single node to a given recurring input pattern. The model was trained with an ecologically valid corpus extracted from text books used to teach reading in French primary schools. The same text books were used to estimate word frequency as manipulated in the behavioral experiment. In order to simulate performance in a lexical decision task, the SOM model was combined with an interactive-activation model (IAM) in order to generate response read-out in terms of predicted RTs and errors. The SOM was used to determine the bigram-word weight values in the IAM. ANOVAs performed on the simulated RTs and errors showed exactly the same pattern as the ANOVAs performed on the behavioral data. Most strikingly, the model accurately simulated the word frequency effect and the interaction between frequency and grade in children. That is, the model correctly showed a diminishing effect of word frequency as a function of simulated number of years of learning to read. Furthermore, the non-significant effect of orthographic neighborhood size found in the simulation results is also in line with the pattern found in the behavioral experiment. The good fit between the model and data, in particular the fact that the model captured the main effects of frequency and grade as well as the evolution of frequency effects with age, suggests that a significant part of learning to read words

involves the kind of self-organized learning mechanisms that are implemented in our model. It is important to note that our simulation study allowed us to demonstrate that the weight-adjustment algorithm of the SOM outperforms a simple frequency-adjusted resting level activation mechanism.

Our model was trained to associate specific combinations of letter sequences (coded as sets of contiguous and non-contiguous bigrams) with a unique higher-level category representation (lexical nodes). These lexical category nodes therefore implement a level of whole-word orthographic representations (orthographic word forms). In the earliest phases of learning, a single lexical representation is activated by the different words that are presented to the model. However, as more and more words are presented to the network, the mapping between orthographic input and lexical representations becomes less and less ambiguous as a function of the number of times a word is presented. By the end of simulated grade 1, a stable mapping begins to emerge between the orthographic input and higher-level lexical representations such that a majority of the words that have been presented to the network now correctly activate a distinct lexical representation (one-to-one mapping). However, some words in the grade 1 pool still have an ambiguous one-to-many mapping from lexical representation to input representation. It is these words that generate an error in the simulated performance of children's lexical decision accuracy. What is critical in the simulation results is that the number of errors is shown to depend significantly on word frequency in grade 1, and that this influence of word frequency diminishes as reading experience increases. Basically, the model's performance to high frequency words asymptotes quite quickly, while performance to low frequency words develops more slowly. This is a direct result of the algorithm used to adjust connection strengths in the model as a function of exposure to a given input. As the weights get closer to their maximum value (1.0 in the present simulations), the change in weight becomes smaller such that with unlimited exposure the weight value converges asymptotically to its maximum value. This is a feature shared by many neural networks, with the consequence that weight changes are relatively large in initial phases of learning and gradually diminish with more and more training.

Our self-organizing model of orthographic learning generated a developmental pattern that provided a very good fit with the learning curve found in children between grades 1 and 5. Moreover, this good fit with the empirical data was obtained with a realistic exposure of the model (in terms of number of epochs) to an ecologically valid training corpus. The relative success of the SOM contrasts with the difficulty of back-propagation networks to generate plausible developmental patterns (Hutzler et al., 2004). In contrast to back-propagation models (Seidenberg & McClelland, 1989; Harm & Seidenberg, 1999; 2004; Plaut et al., 1996), there are no hidden units in the SOM. Nevertheless, the resulting map shows certain non-linearities. For example, an informal analysis of the different maps shows that words tend to be topologically organized into similarity neighborhoods, with clusters of words on the map presenting a certain orthographic similarity. In spite of this generality, in some cases words are located far from their expected similarity cluster. Future research will explore the extent to which the topographic structure of SOMs can capture effects of orthographic similarity neighborhoods in visual word recognition. This research should apply more refined measures of orthographic similarity that go beyond the simple N metric.

The idea that a significant part of the learning of orthographic representations proceeds in an unsupervised fashion fits with current accounts of skilled reading that propose a division of labor between orthographic and phonological processes (Harm & Seidenberg, 2004), or others that make a clear distinction between an orthographic and a phonological pathway in skilled reading (Coltheart, Rastle, Perry, Langdon & Ziegler, 2001; Grainger & Ferrand, 1994; Grainger & Ziegler, 2007; Perry, Ziegler & Zorzi, 2007; Zorzi et al., 1998a). For example, Zorzi et al. (1998a) have shown how a simple two-layered associative network using

the supervised delta learning rule can successfully learn regular sublexical spelling-to-sound correspondences but a lexical procedure is needed to learn irregular words. Our unsupervised SOM provides such a mechanism for learning whole-word orthographic representations that can then map onto whole-word phonology via simple associative learning. This whole-word route from orthography-to-phonology would allow accurate reading aloud of irregular words. It is important to note that Li et al. (2004) have successfully applied a SOM to the learning of phonological representations of spoken words. Future research should therefore examine how our orthographic learning map could be connected to a map of whole-word phonological representations as in the Li et al. model.

Finally, the success of our model also provides indirect confirmation of the role of some form of relative-position coding of letter position information during the processing of orthographic information. Support for this type of input coding was provided in our comparison of the simulation results obtained with an open-bigram version of the IAM and the standard slot-based coding scheme version of the IAM (McClelland & Rumelart, 1981). In future simulation work, we will examine whether our SOM can capture some key phenomena observed with adult participants, such as transposed-letter (Perea & Lupker, 2004; Schoonbaert & Grainger, 2004) and relative-position priming (Peressotti & Grainger, 1999; Grainger et al., 2006; van Assche & Grainger, 2006).

## References

- Andrews, S. (1989). Frequency and Neighborhood Effects on Lexical Access - Activation or Search. *Journal of Experimental Psychology-Learning Memory and Cognition*, 15, 802-814.
- Andrews, S. (1992). Frequency and Neighborhood Effects on Lexical Access - Lexical Similarity or Orthographic Redundancy. *Journal of Experimental Psychology-Learning Memory and Cognition*, 18, 234-254.
- Andrews, S. (1997). The effect of orthographic similarity on lexical retrieval: Resolving neighborhood conflicts. *Psychonomic Bulletin & Review*, 4, 439-461.
- Balota, D., & Chumbley, J. (1984). Are lexical decisions a good measure of lexical access? The role of word frequency in the neglected decision stage. *Journal of Experimental Psychology: Human Perception and Performance*, 10, 340-357.
- Balota, D. A., Cortese, M. J., Sergent-Marshall, S. D., Spieler, D. H., & Yap, M. (2004). Visual word recognition of single-syllable words. *Journal of Experimental Psychology: General*, 133, 283-316.
- Bonin P., Chalard M., Méot A., & Fayol M. (2001). Age-of-acquisition and word frequency in the lexical decision task: Further evidence from the French language. *Current Psychology of Cognition*, 20, 401-443.
- Brysbaert, M. (1996) Word frequency affects naming latency in Dutch when age of acquisition is controlled. *European Journal of Cognitive Psychology*, 8(2), 185-193.
- Burani, C., Marcolini, S., & Stella, G. (2002). How early does morpholexical reading develop in readers of a shallow orthography?. *Brain and Language*, 81, 568-586.
- Carreiras, M., Perea, M. & Grainger, J. (1997). Effects of orthographic neighborhood in visual word recognition: Cross-task comparisons. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 23, 857-871.
- Coltheart, M., Davelaar, E., Jonasson, J. F., & Besner, D. (1977). Access to the internal lexicon. In S. Dornic (Ed.), *Attention and Performance VI* (pp. 535-555). Hillsdale, N.J.: Erlbaum.
- Coltheart, M., Rastle, K., Perry, C., Langdon, R. & Ziegler, J. C. (2001). DRC: A dual route cascaded model of visual word recognition and reading aloud. *Psychological Review*, 108, 204-256.
- Connine, C., Mullennix, J., Shernoff, E., & Yelen, J. (1990). Word familiarity and frequency in visual and auditory word recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 16(6), 1084-1096.
- Dehaene, S., Cohen, L., Sigman, M. & Vinckier, F. (2005). The neural code for written words: a proposal. *Trends in Cognitive Sciences*, 9, 335-341.
- Ducrot, S., Lété, B., Sprenger-Charolles, L., Pynte, J., & Billard, C. (2003). The Optimal Viewing Position Effect in Beginning and Dyslexic Readers. *Current Psychology Letters: Behaviour, Brain & Cognition*, 10(1).
- Forster, K.I. (1976). Accessing the mental lexicon. *New Approaches to Language Mechanisms* (pp. 257-287). R.J. Wales & ECT Walker, Editors.
- Forster, K., & Chambers, S. (1973). Lexical access and naming time. *Journal of Verbal Learning & Verbal Behavior*, Vol. 12(6), 627-635.
- Forster, K. I., & Forster, J. C. (2003). DMDX: A Windows display program with millisecond accuracy. *Behavior Research Methods, Instruments & Computers*, 35, 116-124.
- Frith, U., Wimmer, H., & Landerl, K. (1998). Differences in phonological recoding in German- and English-speaking children. *Scientific Studies of Reading*, 2(1), 31-54.
- Ghyselinck, M., Lewis, M. B., & Brysbaert, M. (2004). Age of acquisition and the cumulative-frequency hypothesis: A review of the literature and a new multi-task

- investigation. *Acta Psychologica*, 115, 43-67.
- Goswami, U. (1993). Toward an interactive analogy model of reading development: Decoding vowel graphemes in beginning reading. *Journal of Experimental Child Psychology*, 56(3), 443-475.
- Goswami, U., & Bryant, P. (1990). *Phonological skills and learning to read*. Lawrence Erlbaum Associates, Inc.
- Goswami, U., & Bryant, P. (1992). Rhyme, analogy, and children's reading. *Reading acquisition* (pp. 49-63). Lawrence Erlbaum Associates, Inc.
- Grainger, J. (1990). Word frequency and neighborhood frequency effects in lexical decision and naming. *Journal of Memory and Language*, 29, 228-244.
- Grainger, J. (2008). Cracking the orthographic code: An introduction. *Language and Cognitive Processes*, 23, 1-35.
- Grainger, J. & Ferrand, L. (1994). Phonology and orthography in visual word recognition: Effects of masked homophone primes. *Journal of Memory and Language*, 33, 218-233.
- Grainger, J., Granier, J.P., Farioli, F., Van Assche, E., & van Heuven, W. (2006). Letter position information and printed word perception: The relative-position priming constraint. *Journal of Experimental Psychology: Human Perception and Performance*, 32, 865-884.
- Grainger, J., & Jacobs, A. (1996). Orthographic processing in visual word recognition: A multiple read-out model. *Psychological Review*, 103(3), 518-565.
- Grainger, J., & Jacobs, A. (1998). On localist connectionism and psychological science. *Localist connectionist approaches to human cognition* (pp. 1-38). Lawrence Erlbaum Associates, Publishers.
- Grainger, J., Muneaux, M., Farioli, F., & Ziegler, J. C. (2005). Effects of phonological and orthographic neighbourhood density interact in visual word recognition. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology*, 58 A(6), 981-998.
- Grainger, J., O'Regan, K., Jacobs, A., & Segui, J. (1989). On the role of competing word units in visual word recognition: The neighborhood frequency effect. *Perception and Psychophysics*, 45, 189-195.
- Grainger, J. & Segui, J. (1990). Neighborhood frequency effects in visual word recognition: A comparison of lexical decision and masked identification latencies. *Perception and Psychophysics*, 47, 191-198.
- Grainger, J., & Van Heuven, W. (2003). Modeling Letter Position Coding in Printed Word Perception. *Mental lexicon: "Some words to talk about words"* (pp. 1-23). Nova Science Publishers, Inc.
- Grainger, J., & Whitney, C. (2004). Does the human mind read words as a whole?. *Trends in Cognitive Sciences*, 8(2), 58-59.
- Grainger, J., & Ziegler, J. C. (2007). Cross-code consistency effects in visual word recognition. In E. L. Grigorenko & A. Naples (Eds.), *Single-word reading: Biological and behavioral perspectives* (pp. 129-157). Mahwah, NJ: Lawrence Erlbaum Associates.
- Harm, M., & Seidenberg, M. (1999). Phonology, reading acquisition, and dyslexia: Insights from connectionist models. *Psychological Review*, 106(3), 491-528.
- Harm, M., & Seidenberg, M. (2004). Computing the Meanings of Words in Reading: Cooperative Division of Labor Between Visual and Phonological Processes. *Psychological Review*, 111(3), 662-720.
- Holcomb, P.J., Grainger, J., & O'Rourke, T. (2002). An electrophysiological study of the effects of orthographic neighborhood size on printed word perception. *Journal of Cognitive Neuroscience*, 14, 938-950.

- Hutzler, F., Ziegler, J.C., Perry, C., Wimmer, H., & Zorzi, M. (2004). Do current connectionist learning models account for reading development in different languages? *Cognition*, *91*(3), 273-296.
- Jacobs, A., & Grainger, J. (1994). Models of visual word recognition: Sampling the state of the art. *Journal of Experimental Psychology: Human Perception and Performance*, *20*(6), 1311-1334.
- Johnston, R. A., & Barry, C. (2006). Age of acquisition and lexical processing. *Visual Cognition*, *13*, 789-845.
- Kohonen, T. (1982). Self-organizing formation of topologically correct feature maps. *Biological Cybernetics*, *43*, 59-69.
- Kohonen, T. (1995). *Self-Organizing Maps*. Springer-Verlag, Berlin-Heidelberg-New York
- Lagus, K., Kaski, S., & Kohonen, T. (2004). Mining massive document collections by the WEBSOM method. *Information Sciences*, *163*(1-3), 135-156.
- Laxon, V., Coltheart, V., & Keating, C. (1988). Children find friendly words friendly too: Words with many orthographic neighbours are easier to read and spell. *British Journal of Educational Psychology*, *58*(1), 103-119.
- Laxon, V., Gallagher, A., & Masterson, J. (2002). The effects of familiarity, orthographic neighbourhood density, letter-length and graphemic complexity on children's reading accuracy. *British Journal of Psychology*, *93*(2), 269-287.
- Lefavrais, P. (1965). *Test de l'Alouette*. Paris: ECPA.
- Lété, B., Sprenger-Charolles, L., & Colé, P. (2004). Manulex: A grade-level lexical database from French elementary-school readers. *Behavior Research Methods, Instruments, & Computers*, *36*, 156-166.
- Li, P., Farkas, I., & MacWhinney, B. (2004). Early lexical development in a self-organizing neural network. *Neural Networks*, *17*(8), 1345-1362.
- McClelland, J., & Rumelhart, D. (1981). An interactive activation model of context effects in letter perception: I. An account of basic findings. *Psychological Review*, *88*(5), 375-407.
- McCloskey, M., & Cohen, N. J. (1989). Catastrophic interference in connectionist networks: The sequential learning problem. In G. H. Bower (Ed.), *The psychology of learning and motivation*, vol. 24 (pp. 109-164). San Diego, CA: Academic Press.
- Miikkulainen, R. (1990). Script Recognition with Hierarchical Feature Maps. *Connection Science*, *2*(1-2), 83-101.
- Miikkulainen, R., Bednar, J. A., Choe, Y., & Sirosh, J. (2005). *Computational Maps in the Visual Cortex*. New York: Springer.
- Monsell, S. (1991). The nature and locus of word frequency effects in reading. *Basic processes in reading: Visual word recognition* (pp. 148-197). Lawrence Erlbaum Associates, Inc.
- Morrison, C.M. & Ellis, A.W. (1995). The roles of word frequency and age of acquisition in word naming and lexical decision. *Journal of Experimental Psychology: Learning, Memory & Cognition*, *21*, 116-133.
- Morton, J. (1969). Interaction of information in word recognition. *Psychological Review*, *76*(2), 165-178.
- Mulatti, C., Reynolds, M.G. & Besner, D. (2006). Neighborhood effects in reading aloud: New findings and new challenges for computational models. *Journal Experimental Psychology: Human Perception and Performance*, *32*(4), 799-810.
- Nagy, W., Herman, P.A., & Anderson, R.C. (1985). Learning words from context. *Reading Research Quarterly*, *20*(2), 233-253.

- Norris, D. (2006). The Bayesian reader: explaining word recognition as an optimal Bayesian decision process. *Psychological Review*, 113(2), 327-357.
- Paap, K., Newsome, S., McDonald, J., & Schvaneveldt, R. (1982). An activation-verification model for letter and word recognition: The word-superiority effect. *Psychological Review*, 89(5), 573-594.
- Perea, M., & Lupker, S. (2004). Can CANISO activate CASINO! Transposed-letter similarity effects with nonadjacent letter positions. *Journal of Memory and Language*, 51(2), 231-246.
- Peressotti, F. & Grainger, J. (1999). The role of letter identity and letter position in orthographic priming. *Perception & Psychophysics*, 61, 691-706.
- Perry, C., Ziegler, J. C., & Zorzi, M. (2007). Nested incremental modeling in the development of computational theories: the CDP+ model of reading aloud. *Psychological Review*, 114, 273-315.
- Plaut, D., McClelland, J., Seidenberg, M., & Patterson, K. (1996). Understanding normal and impaired word reading: Computational principles in quasi-regular domains. *Psychological Review*, 103(1), 56-115.
- Rott, S. (1999). The effect of exposure frequency on intermediate language learners' incidental vocabulary acquisition and retention through reading. *Studies in Second Language Acquisition*, 21, 589-619.
- Rubenstein, H., Garfield, L., & Millikan, J. (1970). Homographic entries in the internal lexicon. *Journal of Verbal Learning & Verbal Behavior*, Vol. 9(5), 487-494.
- Saragi, T., Nation, P. & Meister, G.F. (1978). Vocabulary learning and reading. *System*, 6, 72-78.
- Schoonbaert, S., & Grainger, J. (2004). Letter position coding in printed word perception: Effects of repeated and transposed letters. *Language and Cognitive Processes*, 19(3), 333-367.
- Segui, J. & Grainger, J. (1990). Priming word recognition with orthographic neighbors: Effects of relative prime-target frequency. *Journal of Experimental Psychology: Human Perception and Performance*, 16, 65-76.
- Seidenberg, M., & McClelland, J. (1989). A distributed, developmental model of word recognition and naming. *Psychological Review*, 96(4), 523-568.
- Share, D. L. (1995). Phonological recoding and self-teaching: Sine qua non of reading acquisition. *Cognition*, 55(2), 151-218.
- Sprenger-Charolles, L., Siegel, L. S., & Bonnet, P. (1998). Phonological mediation and orthographic factors in reading and spelling. *Journal of Experimental Child Psychology*, 68, 134-155.
- Stevens, M. & Grainger, J. (2003). Letter visibility and the viewing position effect in visual word recognition. *Perception & Psychophysics*, 65, 133-151.
- Treiman, R., Goswami, U., & Bruck, M. (1990). Not all words are alike: Implications for reading development and theory. *Memory & Cognition*, 18, 559-567.
- Van Assche, E. & Grainger, J. (2006). A study of relative-position priming with superset primes. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 32(2), 399-415.
- Zevin, J., & Seidenberg, M. (2002). Age of acquisition effects in word reading and other tasks. *Journal of Memory and Language*, 47(1), 1-29.
- Ziegler, J. C., & Perry, C. (1998). No more problems in Coltheart's neighborhood: Resolving

- neighborhood conflicts in the lexical decision task. *Cognition*, 68(2), 53-62.
- Zorzi, M., Houghton, G., & Butterworth, B. (1998a). Two routes or one in reading aloud? A connectionist dual-process model. *Journal of Experimental Psychology: Human Perception and Performance*, 24(4), 1131-1161.
- Zorzi, M., Houghton, G., & Butterworth, B. (1998b). The development of spelling-sound relationships in a model of phonological reading. *Language and Cognitive Processes*, 13, 337-371.
- Zorzi, M., Perry, C., Ziegler, J. C., & Coltheart, M. (1999). Category-specific deficits in a self-organizing model of the lexical-semantic system. In D. Heinke, G.W. Humphreys, and A. Olson, editors, *Connectionist models in cognitive neuroscience* (pp 137-148). London: Springer.

Author note

The authors thank Marc Brysbaert and two anonymous reviewers for their feedback on an earlier version of this work. The work was supported by Grant ANR-06-BLAN-0337 (Agence Nationale de la Recherche, France). Address correspondence to Stéphane Dufau, Laboratoire de Psychologie Cognitive, Pôle 3C, Case D, CNRS and Université de Provence, 3 place Victor Hugo, 13331 Marseille Cedex 3, or via e-mail : [stephane.dufau@univ-provence.fr](mailto:stephane.dufau@univ-provence.fr).

## Footnotes

1. The weighting of bigram activation by empirical letter visibilities was motivated for extended tests of the SOM that are not reported in the present study, and is not critical for the results of the simulations presented here.
2. The training regime used in the present simulation study is arguably much closer to the real-life exposure to print of children learning to read, compared with the training regimes that are typically used with back-propagation networks.
3. Here we assume that beginning readers perform the lexical decision task as a word identification task – respond “yes” when a word is recognized, and “no” otherwise.