Do current connectionist learning models account for reading development in different languages?

Florian Hutzler[a], Johannes C. Ziegler[b,c,*], Conrad Perry[d,e], Heinz Wimmer[a], Marco Zorzi[f]

[a]*Universität Salzburg, Salzburg, Austria*
[b]*CNRS, Marseille, France*
[c]*Université de Provence, Aix-en-Provence, France*
[d]*The University of Hong Kong, Hong Kong*
[e]*Macquarie Centre for Cognitive Science, Macquarie University, Sydney, Australia*
[f]*Università di Padova, Padova, Italy*

Abstract

Learning to read a relatively irregular orthography, such as English, is harder and takes longer than learning to read a relatively regular orthography, such as German. At the end of grade 1, the difference in reading performance on a simple set of words and nonwords is quite dramatic. Whereas children using regular orthographies are already close to ceiling, English children read only about 40% of the words and nonwords correctly. It takes almost 4 years for English children to come close to the reading level of their German peers. In the present study, we investigated to what extent recent connectionist learning models are capable of simulating this cross-language learning rate effect as measured by nonword decoding accuracy. We implemented German and English versions of two major connectionist reading models, Plaut et al.'s (Plaut, D. C., McClelland, J. L., Seidenberg, M. S., & Patterson, K. (1996). Understanding normal and impaired word reading: computational principles in quasi-regular domains. *Psychological Review*, *103*, 56–115) parallel distributed model and Zorzi et al.'s (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*, 1131–1161); two-layer associative network. While both models predicted an overall advantage for the more regular orthography (i.e. German over English), they failed to predict that the difference between children learning to read regular versus irregular orthographies is larger earlier on. Further investigations showed that the two-layer network could be brought to simulate the cross-language learning rate effect when cross-language differences in teaching methods (phonics versus whole-word approach) were taken into account. The present work thus shows that in order to

---

adequately capture the pattern of reading acquisition displayed by children, current connectionist models must not only be sensitive to the statistical structure of spelling-to-sound relations but also to the way reading is taught in different countries.

1. Introduction

Writing systems differ in spelling-to-sound consistency.[1] This has a dramatic effect on the speed at which reading skills are acquired. For example, Italian, Spanish, Greek, and Finnish have regular orthographies, in which letters or letter clusters consistently map onto phonemes. At the end of grade 1, children in these countries are typically close to ceiling in terms of reading accuracy (Goswami, Gombert, & Fraca de Barrera, 1998; Seymour, Aro, & Erskine, 2003). In comparison, children learning to read English are faced with a large amount of inconsistency, where the same orthographic patterns can often be pronounced in multiple ways and the same pronunciations can almost always be spelled in multiple ways (e.g. Perry, Ziegler, & Coltheart, 2002; Ziegler, Stone, & Jacobs, 1997). Not surprisingly, it takes children in English-speaking countries much longer to obtain a high level of reading performance compared to children learning more regular orthographies (Goswami et al., 1998; Seymour et al., 2003).

One of the most critical skills for successful reading acquisition is phonological decoding (Share, 1995). Phonological decoding can be accurately measured by examining children's nonword reading performance. Nonword decoding is a crucial skill because it allows children to make the connection between novel letter sequences and words that are already stored in their phonological (spoken word) lexicons. It is this ability to generalize that allows the child to successfully decode and construct orthographic entries for thousands of new words during their first years of education (Share, 1995).

Studies of nonword reading skills show that the acquisition of phonological recoding skills in English is slow and difficult. Mean error rates for nonword reading at the end of grade 1 typically range from 40% to 80% (e.g. Jorm, Share, MacLean, & Matthews, 1984; Juel, Griffith, & Gough, 1986; Seymour et al., 2003; Treiman, Goswami, & Bruck, 1990). In contrast, in Greek, a regular orthography, children of the same age made only about 10% errors when reading words and nonwords (Porpodas, 1999). In a recent review, Landerl (2000) reports that children in regular orthographies like Dutch, German, Greek, Italian, Portuguese or Turkish make no more than 25% errors on nonword reading at the end of grade 1.

---

[1] We use the term *consistency* in a general way to mean consistency in the statistical mapping between orthography and phonology (e.g. Jared, 2002; Kessler & Treiman, 2001; Treiman, Mullennix, Bijeljac-Babic, & Richmond-Welty, 1995). Note that our use of this concept is not restricted to the mapping between bodies and rimes. We use the term *regularity* in a more restricted way to refer to the regularity of grapheme–phoneme correspondences.

Apart from monolingual studies, there have also been some direct cross-language comparisons. A Turkish–English (Oney & Goldman, 1984), an Italian–English (Thorstad, 1991), and a Greek–English (Goswami, Porpodas, & Weelwright, 1997) comparison all replicated superior nonword reading skills for children learning to read regular orthographies. Furthermore, using nonwords derived from number words (by exchanging onsets), Wimmer and Goswami (1994) and Jansen (1995) overcame the methodological problem of insufficient comparability that can arise when stimuli of different orthographies are compared. Interestingly, they both found that there were essentially no differences between children's ability to read the highly frequent number words in the different orthographies, but there were big differences between children's ability to read nonwords. Thus, again, these results suggest that the main problem of the English children lies in their relatively poor phonological decoding skills.

One of the most interesting cross-language comparisons is between German and English. Due to their common Germanic origin, both languages have a very similar orthography and phonology but differ with respect to spelling-to-sound regularity (see Ziegler, Perry, & Coltheart, 2000). This is nicely illustrated by the large number of words that have identical orthographic forms in both languages (*land*, *bank*, *ball*, *zoo*, etc.). This property made it possible to investigate reading development and skilled reading in different orthographies using literally identical stimulus material (Frith, Wimmer, & Landerl, 1998; Landerl, Wimmer, & Frith, 1997; Ziegler, Perry, Jacobs, & Braun, 2001). These studies show a very similar picture to the one summarized above, namely that English children show poorer nonword reading skills compared to German children even when identical stimulus material is used. One prototypical data pattern, taken from the study by Frith et al. (1998), is presented in Fig. 1.
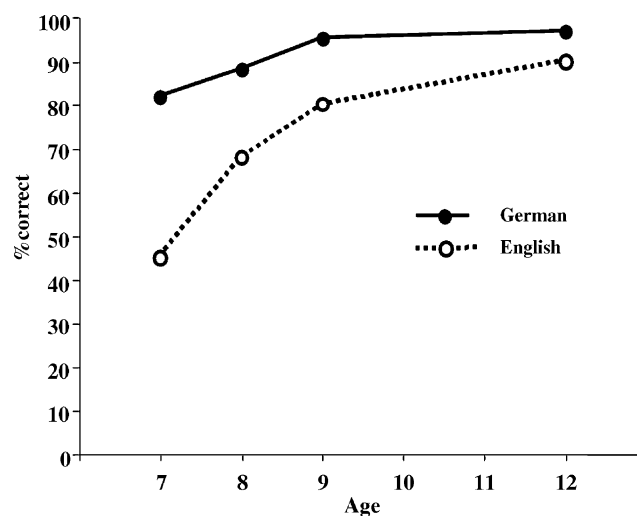

Fig. 1. Prototypical data pattern illustrating the learning rate effect that can be observed in literally every cross-language comparison involving an orthographically consistent (e.g. German) and the relatively less consistent English orthography. Data reproduced from Frith et al. (1998).

In the Frith et al. (1998) study, the authors collected nonword reading accuracy data for 7-, 8-, 9-, and 12-year-old children in Austria and England. The data clearly show that the biggest difference in nonword reading between the German and the English children is early on. As can be seen, in the regular orthography (German), children show much better performance, with reading accuracy levels already above 75% by the age of 7. It takes the English children several years of instruction to catch up with the German children. However, even by the age of 12, phonological decoding of the English children is still less accurate than that of the German children. We will refer to the developmental pattern illustrated in Fig. 1 as the cross-language *learning rate effect*. It is important to note that this effect does not depend on the way errors are coded. This is because in the Frith et al. study, as well as in all of the above-mentioned studies, the errors of the English children were scored in a lenient way, in which every response that was phonologically plausible (even if it was incorrect according to common spelling-to-sound rules) was counted as correct.

In the present research, we were interested in investigating to what extent current connectionist learning models were able to capture cross-language learning. Indeed, a number of connectionist models are potentially capable of simulating developmental data (Harm & Seidenberg, 1999; Plaut, McClelland, Seidenberg, & Patterson, 1996; Seidenberg & McClelland, 1989; Zorzi, Houghton, & Butterworth, 1998a,b). Other models exist for skilled reading (e.g. Coltheart, Rastle, Perry, Langdon, & Ziegler, 2001; Jacobs, Rey, Ziegler, & Grainger, 1998), but, because they do not tackle the question of learning, they are not relevant in the context of the present study. More generally, understanding how models come to solve the stability–plasticity dilemma present in any learning situation is certainly a major challenge that deserves close attention. Typically, connectionist learning models are presented several times with a large training corpus of several thousands of words. Using various learning algorithms, they extract statistical relationships between spelling and sound (see Zorzi, in press, for a review). Developmental data have been simulated with these models, by looking at performance on a given target set before training has been completed (e.g. Zorzi et al., 1998b).

The most influential connectionist model of reading is the triangle model proposed by Plaut et al. (1996), which was an update of the parallel distributed processing (PDP) model of Seidenberg and McClelland (1989). Although the full triangle model contains both a phonological and a semantic pathway, in the present work we only focus on the phonological pathway (i.e. orthography-to-phonology mapping) because this is the pathway that is relevant for nonword decoding skills. The network was designed to learn to read monosyllabic words. It uses a three-layer architecture, in which an orthographic layer is connected to a phonological layer via a hidden-unit layer. During each epoch, the network processes each word. The hidden units compute their states based on the active graphemes and the weights on the connections from them. The phoneme units compute their states based on the activation of the hidden units and the corresponding connection weights. The back-propagation algorithm is used to calculate changes of the connection weights in order to reduce the discrepancy between the generated phoneme activations and the correct patterns. The original network was trained to essentially perfect performance on about 3000 English words in the course of 300 learning epochs.

Plaut et al. (1996) showed that the model did a good job of simulating skilled reading performance. For example, it produced the critical frequency by consistency interaction,

which is a hallmark of skilled reading in English. In addition, it did a much better job of nonword reading (generalization performance) than the original Seidenberg and McClelland (1989) PDP model, which had been criticized for performing poorly on that task. Although the model was able to simulate reading delay in developmental dyslexia, it was never actually tested against developmental data from reading acquisition studies.

The goal of the present study was to investigate whether the triangle model could predict developmental data and most importantly whether it could do so for orthographies that differ in terms of spelling-to-sound consistency. Given that Plaut et al. (1996) showed that the consistency of spelling-to-sound relations was a major factor in network learning, there are good theoretical grounds for expecting that learning in a regular orthography should be faster than learning in a less regular orthography. For the present simulation work, we chose the German–English cross-language comparison for a number of reasons. First, the learning rate effect for these languages is extremely well documented and has proved reliable in a number of studies (Frith et al., 1998; Goswami, Ziegler, Dalton, & Schneider, 2001; Landerl et al., 1997; Wimmer & Goswami, 1994). Second, the input and output domains in these languages are extremely similar, making it possible to test the models on the same set of items.

2. Simulation 1: does the triangle model predict the cross-language learning rate effect?

The phonological pathway of the triangle model (Plaut et al., 1996) was implemented in German and English. Both versions of the model were trained on comparable training corpora matched in size and frequency across languages. In order to explore whether the model would acquire differential nonword reading ability across languages (i.e. predict the cross-language learning rate effect), we tested both implementations on an identical set of nonwords during the course of training.

2.1. Method

2.1.1. Network architecture

The original English model was exactly reconstructed as specified by Plaut et al. (1996) in Simulation 1. That is, the feedforward network consisted of three layers, a grapheme layer (105 units), a hidden layer (100 units), and a phoneme layer (61 units). Connections exist only between grapheme units and hidden units and between hidden units and phoneme units. Each grapheme unit sends activation to each hidden unit, and each hidden unit sends activation to each phoneme unit. Initial weights on connections are small random values between $+0.1$ and -0.1. The large number of grapheme and phoneme units results from a scheme that codes a consonant differently as a function of its appearance before the vowel (i.e. in the onset) or after the vowel (i.e. in the coda). All remaining characteristics of the network, such as activation function, decay, and training procedure (i.e. changes on connection weights based on back-propagation as learning algorithm and cross-entropy as error measure, the global learning rate, the connection

specific learning rate and the momentum) were the same as described by Plaut et al. (1996) in Simulation 1.

The German network had the same architecture as the English network, but used only 82 grapheme and 52 phoneme units. The smaller number of German grapheme units results solely from the onset and coda set. In the vowel set, more units were used for German than for English. One reason for that is that the German vowel set includes quite a few letters with diacritics (e.g. ä), which indicate a change in the pronunciation of the letter. Note that simple differences in the number of grapheme and phoneme units cannot explain differences in the speed of learning because, as the rich literature on the XOR problem convincingly demonstrated (e.g. PDP books), it is mainly the *non-linearity* of the relationships that matters rather than the number of input and output relationships. Given the smaller number of input and output units it seemed appropriate to reduce the number of hidden units to 80 as compared to 100 in the English implementation. Note that this change actually makes very little difference, and using 100 and not 80 hidden units leads to an almost identical pattern of performance. The German grapheme units also include two disambiguation units (TS and SK) and the unit ^ to code whether the word is capitalized or not. The phoneme units include the disambiguation unit /sk/.

2.1.2. Training corpora

For the German corpus, we selected all monosyllabic and monomorphemic words from the CELEX database (Baayen, Piepenbrock, & von Rijn, 1993). All proper nouns and geographical terms were excluded. We also excluded loan words. These were defined as either not being included in the authoritative German dictionary *Duden* (1980) or were marked in the *Duden* as loan words. Examples of such loan words are *Boy*, *Gag* and *chic*. Note that these exclusion criteria were much the same as those applied by Plaut et al., who followed Seidenberg and McClelland (1989), except they excluded loan words by hand. No homographic homophones were included.

The resulting German corpus consisted of only 1293 words compared to about 3000 words in the original English corpus. Reasons for the smaller number of monosyllabic words in German are that the infinitive form of German verbs always consists of two syllables (e.g. *singen* – *to sing*) and that the final -e in many words is pronounced in German but not in English (e.g. *Nase* – *nose*). The selected German monosyllabic words have on average of 4.5 letters (range: 2–8).

To keep the number of words in the German and the English training corpora the same, we randomly reduced the number of English words to 1293, but kept within the reduced English corpus the 13 homographs and the words that the Glushko (1979) pseudowords were derived from. The frequency of the words in the resulting English training corpus was slightly lower than the frequency of the German words (log frequency of 0.24 versus 0.29, respectively). Because log frequency is used to scale weight changes after learning, we reduced the German log frequency by a factor down to 0.24 to provide equal learning prerequisites.

2.1.3. Training procedure

During each sweep through the training corpus (learning epoch), the network processed each word. Hidden units computed their states based on the active graphemes and their

corresponding connection weights. Phoneme units computed their states based on those of the hidden units and their corresponding connection weights. After each word was processed in this way during a learning epoch, back-propagation was used to calculate changes of the connection weights in order to reduce the discrepancy between the generated phoneme activations and the correct pattern. The German model was trained in exactly the same way as the English model.

In order to evaluate the reading performance of the network, Plaut et al. used the following procedure. A phoneme unit within the onset and coda set was considered to be produced by the network when the activity level of the unit was above 0.5 (range −1.0 to +1.0). Within the vowel set, the vowel with the highest activity (even below 0.5) was considered to be produced. When using these criteria, Plaut et al. found that the network correctly pronounced all of the 2972 nonhomographic words of the training corpus after 300 training epochs. For each of the 13 homographs, one of the correct pronunciations was produced. On the consistent and inconsistent pseudowords of Glushko (1979), it produced 98% and 72% correct pronunciations, respectively, which is quite similar to the performance levels of human readers.

Although our implementation of the original English network was only trained on 1293 words, it produced a virtually identical performance as the original Plaut et al. network. After 300 training epochs, it produced only four erroneous pronunciations on nonhomographic words (BEEN as /bin/, ONE as /wOn/, OUR as /Or/ and OWN as /Wn/) and produced a correct reading for each of the 13 homographs. On the Glushko (1979) items, it was correct on 95% and 76% of the consistent and inconsistent pseudowords, respectively.

The implementation of the German network produced only two erroneous pronunciations after 300 epochs of training (i.e. PAPST [pope] and WEG [way]). In the case of PAPST, the only error in the network's pronunciation was a short vowel instead of the correct long vowel. This is of interest, because PAPST is one of the few irregular words of the German orthography (since a consonant cluster after a vowel normally marks this vowel to be pronounced short) – the network's wrong pronunciation is therefore a regularization error. In the case of the nonhomophonic homograph pair *Weg/weg*, only one correct pronunciation was produced.

2.1.4. Testing procedure

Because the critical test is with regard to nonword reading (network generalization), the German and English implementations were tested on a set of 80 nonwords during the course of training. The 80 nonwords were literally identical across the two languages (fot–fot, lank–lank, plock–plock, etc.). All of them were monosyllabic and either three, four, five or six letters long (always 20 items per category). Furthermore, the English and German nonwords were matched in terms of number of letters, body neighborhood (Ziegler & Perry, 1998), letter neighborhood (Coltheart, Davelaar, Jonasson, & Besner, 1977), and consistency ratio (Ziegler et al., 1997). We did not use the Frith et al. (1998) nonwords because they contained multisyllabic items, and their monosyllabic set was not sufficiently matched for a variety of factors that are likely to affect model performance, like word length and neighborhood.

The testing procedure was straightforward. After each sweep through the entire lexicon (1293 words), the 80 nonwords were presented to the model and the number of correct

responses was established. It is important to note that, for the English model, all phono-logically plausible responses were considered correct, even if they did not respect the most frequent grapheme–phoneme correspondences (e.g. *voop* would be considered correct if the /u/ phoneme was either long or short). Thus, as in the human studies (e.g. Frith et al., 1998), a lenient criterion of error coding was adopted for the English model.

2.2. Results

The first non-trivial result that is worth pointing out is that both German and English networks were able to learn the task of word and nonword reading. The models' performance on the critical set of nonwords is presented in Fig. 2. Initially during learning (until about 100 cycles), the learning curves of the English and the German models are close together, with slightly higher performance for the English model. While performance of the English model flattens out after 100 cycles at around 70% correct for nonword reading, performance of the German model keeps on increasing towards asymptotic performance of around 90% after about 200 learning cycles.

When this pattern is compared with the cross-language learning rate effect illustrated in Fig. 1, an interesting discrepancy becomes apparent. While the learning rate effect is best characterized by big differences in early learning phases and small differences in later learning phases, the simulations show the opposite pattern, that is, small differences in early learning phases and big differences in later learning phases.
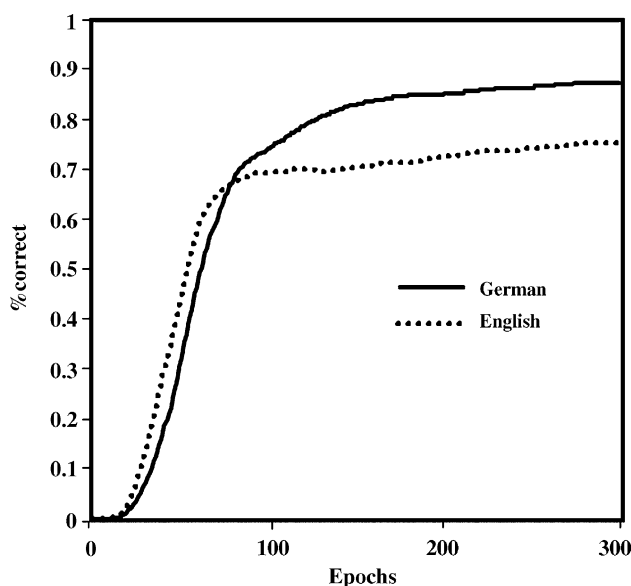

Fig. 2. Nonword reading performance of the German and English implementation of Plaut et al.'s (1996) triangle model during the course of training.

2.3. Discussion

Although both implementations of the triangle model show overall good generalization performance when reading nonwords, they fail to predict the precise direction of the cross-language learning rate effect. That is, the models predict that the higher degree of regularity of the more regular orthography has its main effect later in learning. However, the empirical pattern goes in the opposite way, with an advantage of the more regular language during early learning phases.

What might be responsible for the model's failure to capture the cross-language learning rate effect? One reason might be related to the three-layer architecture of the model. That is, hidden layers in combination with non-linear activation rules are known to pick up higher-order relationships (i.e. allow the learning of non-linear relationships; for a comprehensive illustration see Hinton's (1989) example of object identification). In contrast, the learning rate effect might come about because beginning readers are able to exploit linear orthography–phonology relationships, that is, they might exploit statistical regularities between graphemes and phonemes directly. Moreover, in a three-layer network, initial learning involves mapping a relatively complex set of letter patterns onto the hidden layer (i.e. distributing the orthographic regularities amongst the hidden units) instead of directly strengthening connections between the orthographic and phonological units. This process does not differ much between German and English because both languages have a very similar orthographic structure (i.e. similar orthographic regularities). What differs between the languages is mapping the orthographic regularities onto phonology (i.e. spelling-to-sound consistency). However, in the model, the advantage of the regular over the irregular orthography might only be able to come out once the hidden layer has become fairly stable, that is, during later learning phases.

3. Simulation 2: does a two-layer associative model predict the learning rate effect?

If the reason for the failure of the triangle model to simulate the cross-language learning rate effect is indeed due to the three-layer architecture and the back-propagation learning algorithm, then a two-layer network with a direct mapping between orthography and phonology and a simple associative learning algorithm might do a better job of simulating the effect. In fact, the nonlexical route of Zorzi et al.'s (1998a) dual process model uses a two-layer network and delta-rule learning (i.e. a simple associative algorithm) to learn a direct orthography–phonology mapping. The delta-rule learning procedure is formally equivalent to a classical conditioning law (the Rescorla–Wagner rule; Sutton & Barto, 1981), and has been directly applied to human learning by a number of authors (see, e.g. Gluck & Bower, 1988a,b; Shanks, 1991; Siegel & Allan, 1996, for review). Its use in the present context can thus be supported by appeal to its much wider applicability in predicting learning data. In effect, the model has been successfully used to simulate the development of phonological reading (Zorzi et al., 1998b). Moreover, the same architecture and learning algorithm have been recently used to model the sound-to-spelling mapping in writing (Houghton & Zorzi, 2003).

Although the dual process model of Zorzi et al. contains both a nonlexical and a lexical route, in the present work we only focus on the nonlexical route because this is the pathway that is relevant for nonword decoding skills (this is analogous to using the orthography–phonology pathway of Plaut et al. (1996) as opposed to the full triangle model). Therefore, we implemented a German and English version of the two-layer associative model (see Zorzi et al., 1998a,b, for further details about the English model, and Perry & Ziegler, 2002, for further details about the German model). The prediction was straightforward: if the failure of the triangle model to simulate the cross-language learning rate effect is indeed due to its three-layer architecture (and learning procedure), then Zorzi et al.'s two-layer network should have a better chance of simulating the effect. Model training and testing was done on the same word and nonword sets that were used in Simulation 1.

3.1. Method

3.1.1. Network architecture
The input to the model is a representation of the spelling of a monosyllabic word. Letters in words are represented using a positional code, where each node represents both a letter and the position in the word occupied by that letter. There are no nodes representing combinations of letters, such as graphemes (e.g. TH, EE, etc.). The letter positions are defined with respect to orthographic onset and rime. All letters before the first vowel letter form the onset, and all letters from the vowel onward form the rime. There are three onset positions, and five rime positions. Each letter has a representation (node) at each position, for a total of 208 input nodes. Within each group, successive letters occupy successive positions (i.e. are "left-justified"). Thus, using '*' to denote an empty position, *milk* would be represented as M**ILK**, *old* as ***OLD***, and *strength* as STRENGTH.

The phonological representation has a similar format, with phonemes in a syllable aligned to phonological onset and rime positions in the same way. In this case, there are three onset positions and four rime positions (e.g. /b/ /l/ * /V/ /d/ * *). The phonemic representation recognizes 44 different phonemes of English. All 44 phoneme nodes occur in all seven positions giving 308 output units. The input and output layers are fully connected.

3.1.2. Training and testing procedure
The English and German training corpora were identical to the previous simulation. The models were trained using the delta rule (Widrow & Hoff, 1960). For each spelling–sound pair in the training set, an appropriate orthographic input is established, setting each activated letter-position node to a value of 1. Activations propagate to the output layer, using the dot product net input rule to calculate the inputs to each phoneme unit. Weights are then updated and the next word presented. Note that this is slightly different to the Plaut et al. (1996) model, where weights are updated after all words in the training set have been presented. Connection weights are all initialized to zero, and units have no bias term. Phonemic activations are a sigmoidal function f of their net input, bounding phoneme activations in the range [0,1], and with $f(0) = 0$ (no input, no output). This output activation is compared with the target activation (nodes that should be on have a target

activation of 1, nodes that should be off a target of 0). The error for each phoneme unit is the difference between the target and actual activations. Where errors occur, weights to the offending units are changed according to the delta rule.

The two-layer network model is inherently incapable of learning the whole training set due to the fact that it can only learn linear relationships. Therefore, it cannot be correct for the vast majority of irregular and inconsistent words (e.g. *pint*). This is so because it can only capture the most frequent and consistent (i.e. linear) orthography–phonology mappings. Therefore, the model cannot be trained until the error rate reaches zero. Instead, it is typically trained until errors have apparently reached the global minimum. At that point, the model produces the correct pronunciation of about 81% of the English monosyllabic words and virtually all errors consist in regularizations of the exception words (Zorzi et al., 1998a,b). Using our (reduced) training corpus, the English model produces the correct pronunciation of about 66.5% of the words. The German version produces the correct pronunciation of about 86% of the words. Note that it is explicitly *not* required that the mechanism should be able to correctly read exception words. This is assumed to be achieved through a mediated mapping, which can be based on lexical nodes (as in traditional dual route models; e.g. Houghton & Zorzi, 2003), or on a distributed lexicon (Zorzi et al., 1998a). The aim of the two-layer associative mechanism is simply for it to achieve human-like performance on the *phonological* reading of monosyllabic words and nonwords. In the present simulations, we trained the network for 125,000 word presentations. Note that we use individual word presentations rather than epochs when describing the behavior of this model, because the model learns much quicker than that of Plaut et al. (1996). Thus, if we looked at the simulation results only after every epoch, we would obtain very few data points.

The recall process is the same as that which generates the network's output during training except that a competitive process is implemented at the output layer, whereby multiple candidates compete via lateral inhibition to be the dominant response in a given phoneme position. That is, for a given orthographic input it is possible for more than one phoneme to become active in a given position. Activated phonemes compete via lateral inhibition to become the dominant response. An executable phonological specification is considered to be achieved when all phonemes in each position are either above a response threshold, or are under a "no-response" threshold.

Finally, the testing procedure also examined the generalization performance of the network with the 80 nonwords. However, as mentioned, because the network learns very quickly, we examined performance after the presentation of each 100 words. A similar analysis, based on tracking the network's performance at different points during learning to match it to developmental data, was carried out by Zorzi et al. (1998b) to investigate the development of the sensitivity to VC versus CV constituents in nonword reading (Treiman et al., 1990).

3.2. Results

The results of the two-layer associative network on the critical nonword set are presented in Fig. 3. Both the German and English implementations reach asymptote at about 3000 word presentations with performance levels of about 80% correct
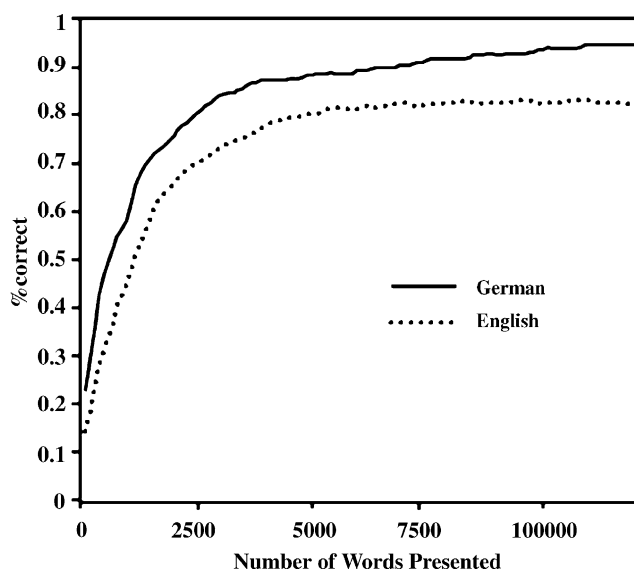
Fig. 3. Nonword reading performance of the German and English implementation of Zorzi et al.'s (1998a) two-layer associative network during the course of training.

nonword readings for the English model and more than 90% correct readings for the German model. There is a consistent 10% advantage of the German network over the English network. This advantage is present right from the beginning of training and remains equally strong until the end of training.

3.3. Discussion

The two-layer associative model predicts a constant advantage of the more regular German orthography over the less regular English orthography. In contrast to the triangle model (Simulation 1), the two-layer associative network predicts an early advantage of German over English, which is characteristic of the cross-language learning rate effect. However, even the two-layer associative network does not fully capture the prototypical pattern illustrated in Fig. 1, because it predicts a constant advantage of the more regular orthography over the less regular orthography, whereas the empirical pattern is best characterized by a large difference (typically around 40%) during early phases of learning and smaller differences during later phases of learning.

Thus, the present simulation shows that the statistical regularity of the orthography–phonology mapping, which is what is learnt by the network, seems to produce a constant advantage of the regular orthography over the irregular orthography. However, some other factor is needed to explain the boost in nonword reading that is obtained for readers of regular orthographies during the initial phases of learning to read (i.e. grade 1). One obvious factor is the teaching method that is used to teach children in the different countries. The more regular languages lend themselves more easily to

grapheme–phoneme teaching (i.e. *phonics*) than the less regular language (i.e. English). This possibility is addressed next.

4. Simulation 3: taking into account teaching methods

Orthographic consistency does not only affect the reliability of the orthography–phonology mapping – a property that is picked up by the two-layer associative model – it also affects the ease with which the mapping can be taught (see Snowling, 1996, for an overview of contemporary teaching approaches). Regular orthographies are most efficiently taught by a pure *phonics* approach, which relies on teaching grapheme–phoneme correspondences. Indeed, this is the (extremely) dominant approach in countries with relatively regular orthographies like Germany, Italy, and Greece (see Landerl, 2000). However, in English-speaking countries, the pure phonics approach does not work as well because inconsistency is maximal for the small grapheme–phoneme mapping (see Treiman et al., 1995). As an example, try to think of a way to teach the word *chalk* using grapheme–phoneme relationships. Therefore, in English, one often finds mixed strategies, where grapheme–phoneme strategies are supplemented by rhyme and whole-word strategies (Goswami & Bryant, 1990; Goswami, Ziegler, Dalton, & Schneider, 2003). Indeed, up to this day, some English-speaking schools still use the whole-word teaching approach (Goodman, 1967; Smith, 1978), in which children are not trained on systematic decoding skills but rather on guessing whole words on the basis of syntactic and semantic redundancies.

The goal of the present simulation was to investigate whether the phonics approach in regular orthographies can explain the large initial advantage of the regular over the irregular orthography, as typically observed in the cross-language comparison between English (irregular orthography) and German (regular orthography). For this purpose, we pre-trained both the German and the English two-layer associative model on a simple set of regular grapheme–phoneme correspondences, thus imitating what might be happening in the course of a pure phonics teaching approach.

4.1. Method

4.1.1. Pre-training

To implement the phonics pre-training, a set of English and German spelling-to-sound correspondences was extracted from commonly used phonics teaching programs. For English, a total of 45 grapheme–phoneme rules were taken from a recent phonics program (*Jolly Phonics*, Lloyd, 1999) and a compilation of rules that appear across eight different phonics programs (Adams, 1990). For German, 54 grapheme–phoneme rules were taken from the three most commonly used phonics teaching programs (Dummer-Smoch & Hackethal, 1994; Eibl, Lampée-Baumgartner, Borries, & Tauscheck, 1996; Krenn & Kowarik, 1988). All the rules that were use are listed in Appendix A.

In order to analyze the consistency characteristics of these rules and to spot potentially consistent grapheme–phoneme correspondences not mentioned in these programs, we extracted all single-letter grapheme–phoneme correspondences, the most frequent

two-letter grapheme–phoneme correspondences, and the most frequent multi-letter graphemes from the German and English version of the CELEX database. For all of these correspondences, a simple consistency ratio (e.g. Treiman et al., 1995) was computed. This analysis showed that all of the English and German single-letter phonics rules ranked amongst the most consistent single-letter correspondences. Concerning the two-letter graphemes, the English phonics programs included 18 rules that ranked amongst the most consistent two-letter rules. However, an additional 11 grapheme–phoneme corres-pondences proved to be highly consistent but were not mentioned in any of the phonics programs. For German, 11 two-letter rules ranked amongst the most consistent two-letter correspondences. Three additional correspondences were highly consistent but they were not mentioned in any of the phonics programs. The phonics programs also missed three highly consistent multi-letter rules for English and two for German. To increase the scope of pre-training, these highly consistent correspondences were added in both languages. In addition, seven rules for consonant doublets were added to both languages. These rules were explicitly listed in the German phonics programs but not in the English programs. To keep the two sets comparable across orthographies, we also added these rules for English. Finally, four phonotactic rules had to be added to the German set in order to take care of consonant devoicing at the ends of words.

This combined selection procedure resulted in a total of 66 pre-training rules for the English orthography and 64 pre-training rules for the German orthography. Since in the two-layer associative model grapheme–phoneme correspondences for consonants have to be taught separately for all different positions in both the onset and the coda (e.g. the occurrence of the letter "p" in the beginning or at the end of a word [e.g. *pool* versus *stop*] has to be taught separately), the final set of computational rules contained 105 rules for German and 109 rules for English. Of course, it might have been possible to construct an even larger number of rules by including low-frequency correspondences. However, rather than providing an exhaustive list of correspondences (e.g. Coltheart, Curtis, Atkins, & Haller, 1993), our main goal was to make rules representative of common phonics programs.

After selecting the grapheme–phoneme relationships, the network was trained for 50 epochs. This meant that all grapheme–phoneme relationships were presented to the network in all of the positions, in which they could occur in the input–output representation of the network. For example, when the English network was given the initial consonant "p" (i.e. p*******), it had to generate the output /p/ (i.e. /p/******). All correspondences in both sets were learned correctly. Note that training the network for different amounts of cycles (e.g. 100 epochs) made very little difference to the pattern of performance.

4.1.2. Training and testing procedure

After the networks had already been given 50 epochs of grapheme–phoneme pre-training, both networks were trained on the 1293 word corpus in exactly the same way as before. One might wonder whether switching from grapheme–phoneme training to word training would cause catastrophic interference (McCloskey & Cohen, 1989). This is not the case, because the same associations learned during grapheme–phoneme pre-training were also "embedded" in the word training set. For example, the association between

grapheme "ea" and phoneme /i/ that is formed during pre-training is not "unlearned" by later training on the word corpus because the same association will be strengthened by spelling–sound pairs such as MEAN → /min/. It is interesting to note that pre-training (although in a different context) has been claimed to be important for modeling learning and development with neural networks, as it enables resistance to catastrophic forgetting (Altmann, 2002).

To test whether the network would exhibit the learning rate effect, the pre-trained models were again presented with the set of 80 nonwords and accuracy was measured after each 100 word representations. Response coding was performed as in the previous simulation.

4.2. Results

The results of the pre-trained versions of the English and German two-layer associative model are presented in Fig. 4A. Inspection of this figure shows that the pre-trained German model exhibited an initial advantage of about 35% over the pre-trained English model. As learning proceeded, this advantage decreased to about 10%. Thus, simulations with the pre-trained models begin to capture the cross-language learning rate effect.

Given what we said earlier about the difficulty of applying a pure phonics approach in the relatively irregular English orthography, the most appropriate comparison might be between the pre-trained German model imitating the phonics approach and the original (not pre-trained) English model imitating the whole-word approach. This comparison is illustrated in Fig. 4B. As can be seen in this figure, the initial advantage of the more regular German orthography over the less regular English orthography increased well above 40%, which is very close to the empirical pattern typically observed (e.g. Frith et al., 1998). The present simulations make it quite clear that the German network benefits rather dramatically from the phonics pre-training regime. One interesting question is whether the English network benefits from pre-training at all. This question can be addressed by comparing the normal model with the pre-trained model within each language. This comparison allows us to estimate to what extent both networks benefit from the phonics pre-training. This comparison is illustrated in Fig. 5A,B.

As can be seen in Fig. 5A, the English network benefits to a small extent from pre-training. These benefits are in the order of 10%, and they are restricted to early learning phases. In contrast, as can be seen in Fig. 5B, the German network greatly benefits from the phonics pre-training regime. The benefits are in the order of 40% in early learning phases and extend in time far beyond the benefits of the English network.

4.3. Discussion

The present simulation showed that the cross-language learning rate effect can be simulated in a two-layer associative network when teaching method is taken into account. This was done by pre-training the model on a simple set of correspondences, imitating what happens during intensive phonics teaching, which is typical of that given to children in many European countries. The results present a striking qualitative fit to
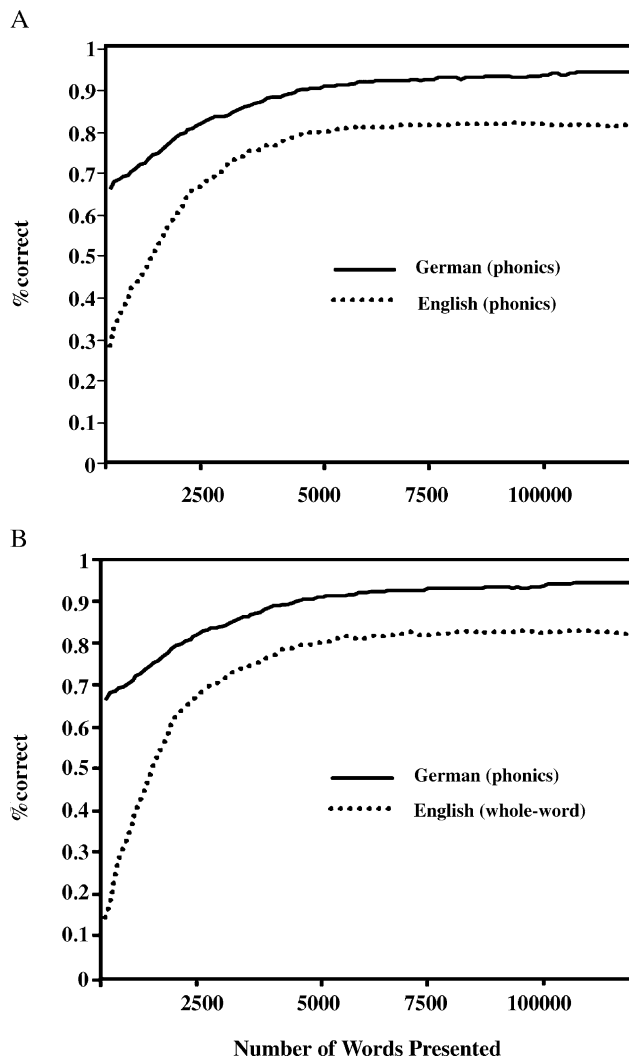
A

B

Number of Words Presented

Fig. 4. Nonword reading performance of Zorzi et al.'s (1998a) two-layer associative network when both implementations are pre-trained using a phonics regime (A) and when the German phonics approach is compared to the English whole-word approach (B).

the data pattern found in a variety of studies comparing reading development in regular orthographies, like German, with reading development in English (e.g. Frith et al., 1998, Fig. 1).

One important question is whether English children can benefit from phonics teaching, and if so, whether they do so to the same extent as the children in more regular orthographies. The simulations suggest that the English network does benefit
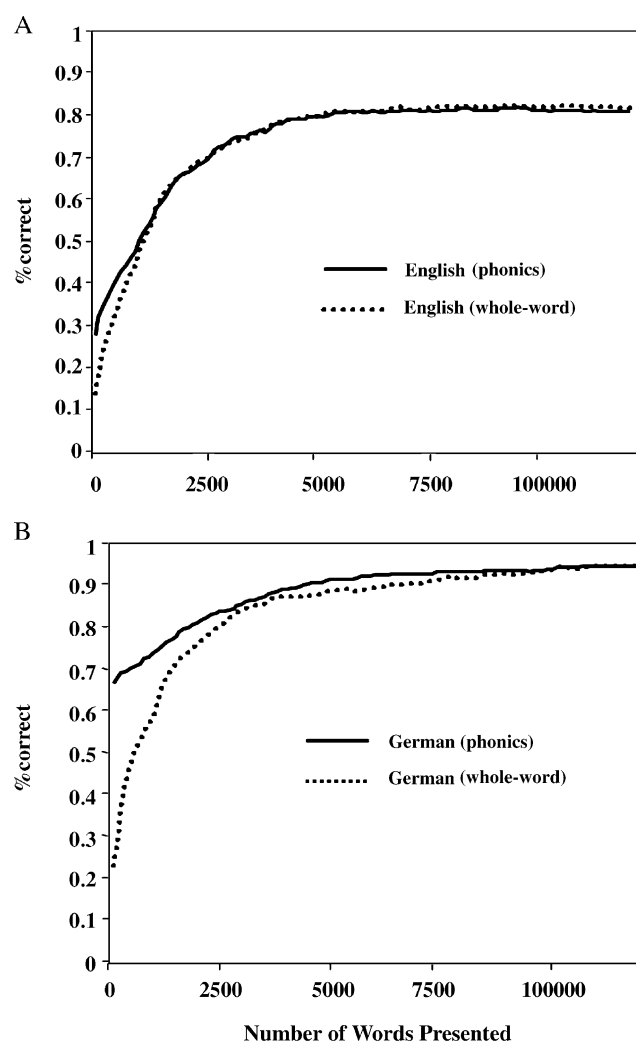
Fig. 5. Benefits due to phonics pre-training of the English and German versions of the two-layer associative network (A,B, respectively).

from the phonics pre-training regime. However, the benefits are smaller and more restricted to early learning phases than those of the German network. This fits well the empirical results reported by Landerl (2000), who found that English children receiving phonics teaching outperformed the English children being taught using the whole-word approach. However, the English children who received the phonics teaching did not reach the same level of performance as children learning to read more regular orthographies.

5. General discussion

Cross-language research over the past decade has shown that learning to read a relatively irregular orthography is harder and takes longer than learning to read a relatively regular orthography (Frith et al., 1998; Goswami et al., 1997, 1998, 2001, 2003). In a recent cross-language study that compared reading acquisition in 13 European countries, Seymour et al. (2003) have shown that English is the hardest European orthography to acquire.

When reading acquisition of English is compared to reading acquisition of a more regular orthography in developmental studies, it is typically observed that the biggest advantage of the more regular orthography is during the first year of reading instruction. At the end of grade 1, for instance, nonword reading performance is about 80% for children in more regular orthographies compared with 40% for the English children (e.g. Frith et al., 1998; Goswami et al., 1997; Seymour et al., 2003). Even when children in different countries are matched on reading age according to standardized tests, as was the case in the German–English comparison by Goswami et al. (2001), the English children with a reading age of about 7 years read only about 30.9% of the nonwords correctly compared to 87.6% being read correctly by the German children. By the age of 12, accuracy of the English children has come closer to that of the German children but a small advantage for the German children remains. This is the pattern reflected in the cross-language learning rate effect illustrated in Fig. 1.

The goal of the present study was to see to what extent current connectionist models were able to capture this cross-language learning rate effect. The focus was on learning models, like Plaut et al.'s (1996) influential triangle model, because only those have the potential to capture a developmental effect. The question was whether the cross-language learning rate effect would emerge in these learning models simply as a consequence of being exposed to large language-specific word corpora or whether additional procedures would be necessary to capture the effect. Because it had been demonstrated that connectionist models are capable of picking up statistical regularities in the orthography–phonology mapping, the prediction was that network performance should benefit when the orthography to be learned is of relatively high consistency.

The results showed that German and English implementations of Plaut's triangle model displayed a better nonword reading performance for the regular German orthography compared to the less regular English orthography. However, the network predicted no cross-language differences during initial learning phases and increasingly larger differences during later learning phases, which is the opposite of the empirical pattern that characterizes the cross-language learning rate effect. We speculated that the three-layer architecture of the model might be responsible for this problem, because in such a network initial learning involves mapping a relatively complex set of letter patterns onto the hidden layer (i.e. distribute the orthographic regularities amongst the hidden units). Because orthographic structure per se does not differ much between German and English, it might be the case that no differences between the two networks are seen in early learning phases. One might argue that it is not until the hidden layer has become stable that spelling-to-sound consistency can affect the mapping from the hidden-layer level to

the phonological output level in a meaningful way (this is more generally known as the "*moving target*" problem; see Ratcliff, 1990).

If the complex architecture of the triangle model is responsible for the failure, then a network with a direct associative orthography–phonology mapping should be better prepared to capture the effect. Indeed a cross-language implementation of a two-layer associative network (Zorzi et al., 1998a,b) was able to predict an initial advantage of German over English. However, even this network did not perfectly capture the learning rate effect because it only predicted a constant advantage of the regular over the irregular orthography, whereas the empirical pattern shows a big initial advantage that is decreasing over the course of learning.

On the basis of these results, we suggest that the German–English difference is best described as an interaction between regularity/consistency of the mapping and teaching method. The constant advantage of the more regular orthography over the less regular orthography is picked up by connectionist networks, because they are sensitive to statistical regularities in the input–output mapping. However, the initially bigger advantage of the more regular over the less regular orthography can only be picked up when teaching method is taken into account. Indeed, when a teaching regime was applied that specifically imitated the phonics approach typically found in regular orthographies (e.g. Landerl, 2000), the two-layer associative network did a good job of predicting the cross-language learning rate effect. We take this as a computational demonstration for the claim that a plausible model of reading development has to be sensitive to both statistical regularities in the input–output mapping as well as constraints imposed by the learning environment.

In sum, we argue that it is probably too simplistic to assume that developmental reading effects simply emerge when connectionist models are exposed to a large word corpus. Instead, in accounting for developmental reading effects, there is a role for how things are being learned. Thus, the present research shows that in understanding learning to read one has to take into account both the statistical structure of what is being learned as well as the particular learning environment.

Acknowledgements

Appendix A. English and German rule sets used for pre-training in Simulation 3

Indices indicate inclusion in specific phonics teaching programs. Phonetic symbols are taken from the CELEX database.

English Grapheme	Phoneme	German Grapheme	Phoneme
Single-letter graphemes			
a	@ [a,b]	a	& [c,d,e]
		ä	) [c,d,e]
b	b [a,b]	b	b [c,d,e]
		b	p [f]
d	d [a,b]	d	d [c,d,e]
		d	t [f]
e	e [a,b]	e	E [c,d,e]
f	f [a,b]	f	f [c,d,e]
g	g [a,b]	g	g [c,d,e]
		g	k [f]
h	h [a,b]	h	h [c,d,e]
i	i [a,b]		
j	j [a,b]	j	j [c,d,e]
k	k [a,b]	k	k [c,d,e]
l	l [a,b]	l	l [c,d,e]
m	m [a,b]	m	m [c,d,e]
n	n [a,b]	n	n [c,d,e]
o	a [a,b]	o	O [c,d,e]
		ö	ʃ [c,d,e]
p	p [a,b]	p	p [c,d,e]
r	r [a,b]	r	r [c,d,e]
s	s [a,b]	s	s [c,d,e]
		s	z [f]
t	t [a,b]	t	t [c,d,e]
u	ʌ [a,b]	u	U [c,d,e]
		ü	Y [c,d,e]
v	v [a,b]	v	f [c,d,e]
w	w [a,b]	w	v [c,d,e]
y	I [a,b]		
z	z [a,b]	z	% [c,d,e]
Two-letter graphemes			
		aa	a [e]
ai	A [a]	ai	W [c]
au	o [a]	au	B [c,d,e]
ay	A [a]		
ch	J [a,b]	ch	x [c,d,e]
ck	k [b]	ck	k [c,e]
		dt	t

(*continued*)

English Grapheme	Phoneme	German Grapheme	Phoneme
ea	E[a]		
ee	E[a,b]	ee	e[e]
ei	A	ei	W[c,d,e]
eu	U	eu	X[c,d,e]
ew	U[a]		
ey	A		
gn	n		
ie	I[a]	ie	i[c,e]
kn	n		
ng	N[a]	ng	N
oa	O[a]		
oe	O		
oi	Y[a]		
		oo	o[e]
ou	W[a]		
ow	W[a]		
oy	Y[a]		
		pf	#[c,e]
ph	f	ph	f
sh	S[a,b]		
th	T[a]		
		ts	%
		tz	%[e]
ue	U[a]		
ui	U		
uy	I		
wh	w		
wr	r		
Multi-letter graphemes and rules			
a*e	A[a,b]		
		ah	a[e]
		äh	)[e]
		auh	B[e]
		eh	e[e]
i*e	I[a,b]		
		ih	i[e]
		ieh	i[e]
o*e	O[a]		

(continued)

English Grapheme	Phoneme	German Grapheme	Phoneme
		oh	o[e]
		öh	I[e]
		sch	S[c,d,e]
tch	J	tch	J
u*e	U[a]		
		uh	u[e]
		üh	y[e]
augh	o		
eigh	A		
tsch	J	tsch	J
Doublets			
ff	f	ff	f[e]
ll	l	ll	l[e]
nn	n	nn	n[e]
pp	p	pp	p[e]
rr	r	rr	r[e]
ss	s	ss	s[e]
tt	t	tt	t[e]

[a]Lloyd (1999); [b]Adams (1990); [c]Eibl et al. (1996); [d]Dummer-Smoch and Hackethal (1994); [e]Krenn and Kowarik (1988); [f]phonotactic rules for final devoicing (German-specific phonotactic constraint).

References

Adams, M. J. (1990). *Beginning to read: thinking and learning about print*. Cambridge, MA: MIT Press.

Altmann, G. T. M. (2002). Learning and development in neural networks – the importance of prior experience. *Cognition*, *85*, B43–B50.

Baayen, R. H., Piepenbrock, R., & van Rijn, H. (1993). *The CELEX lexical database (CD-ROM)*. Philadelphia, PA: Linguistic Data Consortium, University of Pennsylvania.

Coltheart, M., Curtis, B., Atkins, P., & Haller, M. (1993). Models of reading aloud: dual-route and parallel-distributed-processing approaches. *Psychological Review*, *100*, 589–608.

Coltheart, M., Davelaar, E., Jonasson, J. T., & Besner, D. (1977). Access to the internal lexicon. In S. Dornic (Ed.), (*VI*) (pp. 535–555). *Attention and performance*, London: Academic Press.

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.

Duden (1980). *Rechtschreibung der deutschen Sprache und der Fremdwörter*. Mannheim, Wien, Zürich: Bibliographisches Institut.

Dummer-Smoch, L., & Hackethal, R. (1994). *Handbuch zum Kieler Leseaufbau*. Kiel: Veris Verlag.

Eibl, L., Lampée-Baumgartner, T., Borries, W., & Tauscheck, E. (1996). *Mimi die Lesemaus: Arbeitsheft*. Oldenburg: Veritas.

Frith, U., Wimmer, H., & Landerl, K. (1998). Differences in phonological recoding in German- and English-speaking children. *Scientific Studies of Reading*, *2*, 31–54.

Gluck, M. A., & Bower, G. H. (1988a). Evaluating an adaptive network model of human learning. *Journal of Memory and Language*, *27*, 166–195.

Gluck, M. A., & Bower, G. H. (1988b). From conditioning to category learning: an adaptive network model. *Journal of Experimental Psychology: General*, *117*, 227–247.

Glushko, R. J. (1979). The organisation and activation of orthographic knowledge in reading aloud. *Journal of Experimental Psychology: Human Perception and Performance*, *5*, 674–691.

Goodman, K. S. (1967). Reading: a psycholinguistic guessing game. *Journal of the Reading Specialist*, *6*, 126–135.

Goswami, U., & Bryant, P. E. (1990). *Phonological skills and learning to read*. Hillsdale, NJ: Lawrence Erlbaum.

Goswami, U., Gombert, J. E., & Fraca de Barrera, L. (1998). Children's orthographic representations and linguistic transparency: nonsense word reading in English, French, and Spanish. *Applied Psycholinguistics*, *19*, 19–52.

Goswami, U., Porpodas, C., & Weelwright, S. (1997). Children's orthographic representations in English and Greek. *European Journal of Psychology of Education*, *3*, 273–292.

Goswami, U., Ziegler, J. C., Dalton, L., & Schneider, W. (2001). Pseudohomophone effects and phonological recoding procedures in reading development in English and German. *Journal of Memory and Language*, *45*, 648–664.

Goswami, U., Ziegler, J. C., Dalton, L., & Schneider, W. (2003). Nonword reading across orthographies: how flexible is the choice of reading units? *Applied Psycholinguistics*, *24*, 235–247.

Harm, M. W., & Seidenberg, M. S. (1999). Phonology, reading acquisition and dyslexia: insights from connectionist models. *Psychological Review*, *106*, 491–528.

Hinton, G. E. (1989). Connectionist learning systems. *Artificial Intelligence*, *40*, 185–234.

Houghton, G., & Zorzi, M. (2003). Normal and impaired spelling in a connectionist dual-route architecture. *Cognitive Neuropsychology*, *20*, 115–162.

Jacobs, A. M., Rey, A., Ziegler, J. C., & Grainger, J. (1998). MROM-p: an interactive activation, multiple readout model of orthographic and phonological processes in visual word recognition. In J. Grainger, & A. M. Jacobs (Eds.), *Localist connectionist approaches to human cognition* (pp. 147–188). *Scientific psychology series*, Mahwah, NJ: Lawrence Erlbaum Associates.

Jansen, M. E (1995). *Zur Frage der allgemeinen Gültigkeit "englischer" Modelle des Lesenlernens [On the generality of "English" models of learning to read]*. Unpublished master's thesis, University of Salzburg, Salzburg, Austria.

Jared, D. (2002). Spelling-sound consistency and regularity effects in word naming. *Journal of Memory and Language*, *46*, 723–750.

Jorm, A. F., Share, D. L., MacLean, R., & Matthews, R. G. (1984). Phonological recoding skills and learning to read – a longitudinal-study. *Applied Psycholinguistics*, *5*(3), 201–207.

Juel, C., Griffith, P. L., & Gough, P. B. (1986). Acquisition of literacy – a longitudinal-study of children in 1st-grade and 2nd-grade. *Journal of Educational Psychology*, *78*(4), 243–255.

Kessler, B., & Treiman, R. (2001). Relationship between sounds and letters in English monosyllables. *Journal of Memory and Language*, *44*, 592–617.

Krenn, R., & Kowarik, O. (1988). *Horchen – Zeigen – Lesen*. Wien: Jugend und Volk.

Landerl, K. (2000). Influences of orthographic consistency and reading instruction on the development of nonword reading skills. *European Journal of Psychology and Education*, *15*(3), 239–257.

Landerl, K., Wimmer, H., & Frith, U. (1997). The impact of orthographic consistency on dyslexia: a German–English comparison. *Cognition*, *63*, 315–334.

Lloyd, S. (1999). *The phonics handbook* (3rd ed.). London: Jolly Learning.

McCloskey, M., & Cohen, N. (1989). Catastrophic interference in connectionist networks: the sequential learning problem. *The Psychology of Learning and Motivation*, *24*, 109–165.

Oney, B., & Goldman, S. R. (1984). Decoding and comprehension skills in Turkish and English – effects of the regularity of grapheme phoneme correspondences. *Journal of Educational Psychology*, *76*(4), 556–568.

Perry, C., & Ziegler, J. C. (2002). A cross-language computational investigation of the length effect in reading aloud. *Journal of Experimental Psychology: Human Perception and Performance*, *28*, 990–1001.

Perry, C., Ziegler, J. C., & Coltheart, M. (2002). How predictable is spelling? An analysis of sound-spelling contingency in English. *The Quarterly Journal of Experimental Psychology*, *55A*, 897–915.

Plaut, D. C., McClelland, J. L., Seidenberg, M. S., & Patterson, K. (1996). Understanding normal and impaired word reading: computational principles in quasi-regular domains. *Psychological Review*, *103*, 56–115.

Porpodas, C. D. (1999). Patterns of phonological and memory processing in beginning readers and spellers of Greek. *Journal of Learning Disabilities*, *32*(5), 406–416.

Ratcliff, R. (1990). Connectionist models of recognition memory: constraints imposed by learning and forgetting functions. *Psychological Review*, *97*, 285–308.

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

Seymour, P. H. K., Aro, M., & Erskine, J. M. (2003). Foundation literacy acquisition in European orthographies. *British Journal of Psychology*, *94*, 143–174.

Shanks, D. R. (1991). Categorization by a connectionist network. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *17*, 433–443.

Share, D. L. (1995). Phonological recoding and self-teaching: sine qua non of reading acquisition. *Cognition*, *55*, 151–218.

Siegel, S., & Allan, L. G. (1996). The widespread influence of the Rescorla-Wagner model. *Psychonomic Bulletin and Review*, *3*, 314–321.

Smith, F. (1978). *Understanding reading*. New York: Holt, Rinehart, & Winston.

Snowling, M. J. (1996). Annotation: contemporary approaches to the teaching of reading. *Journal of Child Psychology and Psychiatry*, *37*(2), 139–148.

Sutton, R. S., & Barto, A. G. (1981). Toward a modern theory of adaptive networks: expectation and prediction. *Psychological Review*, *88*, 135–170.

Thorstad, G. (1991). The effect of orthography on the acquisition of literacy skills. *British Journal of Psychology*, *82*, 527–537.

Treiman, R., Goswami, U., & Bruck, M. (1990). Not all nonwords are alike – implications for reading development and theory. *Memory and Cognition*, *18*(6), 559–567.

Treiman, R., Mullennix, J., Bijeljac-Babic, R., & Richmond-Welty, E. D. (1995). The special role of rimes in the description, use, and acquisition of English orthography. *Journal of Experimental Psychology: General*, *124*, 107–136.

Widrow, G., & Hoff, M. E. (1960). Adaptive switching circuits. *Institute of Radio Engineers, Western Electronic Show and Convention, Convention Record, Part 4* (pp. 96–104). New York: IRE.

Wimmer, H., & Goswami, U. (1994). The influence of orthographic consistency on reading development – word recognition in English and German children. *Cognition*, *51*(1), 91–103.

Ziegler, J. C., & Perry, C. (1998). No more problems in Coltheart's neighborhood: resolving neighborhood conflicts in the lexical decision task. *Cognition*, *68*, 53–62.

Ziegler, J. C., Perry, C., & Coltheart, M. (2000). The DRC model of visual word recognition and reading aloud: an extension to German. *European Journal of Cognitive Psychology*, *12*, 413–430.

Ziegler, J. C., Perry, C., Jacobs, A. M., & Braun, M. (2001). Identical words are read differently in different languages. *Psychological Science*, *12*, 379–384.

Ziegler, J. C., Stone, G. O., & Jacobs, A. M. (1997). What's the pronunciation for -OUGH and the spelling for /u/? A database for computing feedforward and feedback inconsistency in English. *Behavior Research Methods, Instruments, & Computers*, *29*, 600–618.

Zorzi, M (in press). Computational models of reading. In G. Houghton (Ed.), *Connectionist models in psychology*. London: Psychology Press.

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*, 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.