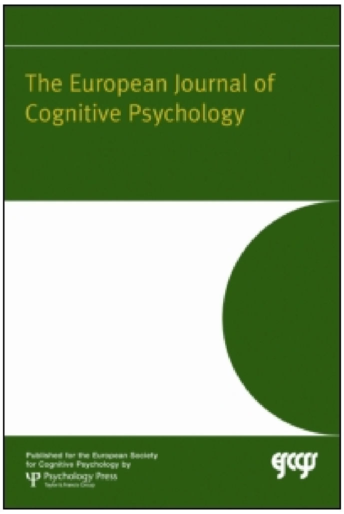

European Journal of Cognitive Psychology
Publication details, including instructions for authors and subscription information:
http://www.informaworld.com/smpp/title~content=t713734596

The connectionist dual process (CDP) approach to modelling reading aloud

Marco Zorzi[a]
[a] Department of General Psychology and Centre for Cognitive Science, University of Padova, Padova, Italy

Online publication date: 21 July 2010

To cite this Article Zorzi, Marco(2010) 'The connectionist dual process (CDP) approach to modelling reading aloud', European Journal of Cognitive Psychology, 22: 5, 836 — 860

To link to this Article: DOI: 10.1080/09541440903435621

URL: http://dx.doi.org/10.1080/09541440903435621

EUROPEAN JOURNAL OF COGNITIVE PSYCHOLOGY
2010, 22 (5), 836–860

Psychology Press
Taylor & Francis Group

The connectionist dual process (CDP) approach to modelling reading aloud

Marco Zorzi

Department of General Psychology and Centre for Cognitive Science, University of Padova, Padova, Italy

This paper reviews the Connectionist Dual Process (CDP) approach to modelling reading aloud, from the computational principles that motivate the model's connectionist dual-route architecture to the most recent developments guided by a nested incremental modelling strategy. New simulations based on a greatly simplified, feedforward version of the model (ffCDP+) demonstrate that the models' success in accounting for key phenomena in word naming relies on the nature of its sublexical route. The results with ffCDP+, where the parameter-heavy, interactive lexical route is turned into a simple frequency-weighted activation of lexical phonology, show that the two-layer associative network of phonological assembly is the core component in CDP+ and highlight the intrinsic modularity of the model. Further developments of the model and some directions for future research are discussed.

Keywords: Reading aloud; Word naming; Computational modelling; Connectionist models; Neural networks.

Visual word recognition and reading aloud is the area of cognitive psychology where computational modelling has probably achieved its greatest success. Seminal studies published in the 1980s (McClelland & Rumelhart, 1981; Rumelhart & McClelland, 1982; Seidenberg & McClelland, 1989) have dramatically changed the way we think about the basic processes in oral reading of single words. More recent modelling work has produced highly detailed simulations of various aspects of the reading process (e.g., Coltheart, Curtis, Atkins, & Haller, 1993; Coltheart, Rastle, Perry, Langdon, & Ziegler, 2001; Grainger & Jacobs, 1996; Harm & Seidenberg, 1999, 2004; Perry, Ziegler, & Zorzi, 2007; Plaut, McClelland, Seidenberg, & Patterson, 1996;

Correspondence should be addressed to Marco Zorzi, Dipartimento di Psicologia Generale, Università di Padova, via Venezia 8, 35131 Padova, Italy. E-mail: marco.zorzi@unipd.it

The CDP Web site can be found at http://ccnl.psy.unipd.it/CDP.html. This study was supported by a grant from the European Research Council (No. 210922-GENMOD).

http://www.psypress.com/ecp DOI: 10.1080/09541440903435621

Zorzi, Houghton, & Butterworth, 1998b) and it has greatly improved our understanding of both normal performance in skilled readers and impaired processing in dyslexic readers (see Zorzi, 2005, for a review).

In this paper I review the computational principles that motivated the Connectionist Dual Process (CDP) model of reading aloud (Zorzi et al., 1998b) and its development over the past 10 years (Perry et al., 2007, in press; Zorzi, 1999, 2000). The first two sections discuss the computational motivation for the CDP approach to learning to read and the features of CDP that make it different from other computational models of reading aloud. The third section reviews the recent developments of CDP and highlight the guiding role of the *nested incremental modelling* strategy (Perry et al., 2007). I discuss how CDP+ (Perry et al., 2007), designed by building upon the strengths of CDP while eliminating its weaknesses, has turned out to be the most successful computational model of reading aloud to date. In the fourth section I present new simulations demonstrating that the success of the CDP+ model is entirely driven by its sublexical component, the two-layer associative network of phonological assembly. In the last two sections I discuss further developments of the model as well as the most important directions for future research.

THE CONNECTIONIST DUAL PROCESS APPROACH TO LEARNING IN QUASIREGULAR DOMAINS

The seminal work of Seidenberg and McClelland (1989) showed that key phenomena in word naming (e.g., frequency, consistency) can emerge in a connectionist network where the mapping between orthographic (input) and phonological (output) representations is learnt from a corpus of spelling–sound exemplars through the error-backpropagation algorithm (Rumelhart, Hinton, & Williams, 1986). Although the model envisaged a second pathway between spelling and sound mediated by semantic representations, its processing assumptions did not conform to the classic distinction between rule-based phonological assembly versus phonological retrieval from a lexical memory that was ubiquitous in dual-route models of reading aloud (e.g., Coltheart, 1978; Patterson & Morton, 1985). This led to the widely held belief that connectionist models had a "single-route" architecture that challenged dual-route theories of the reading system.

The breakthrough that led to the Connectionist Dual Process (CDP) model (Zorzi et al., 1998b) was the discovery that a dual-route processing system can emerge from the interaction of task demands and initial network architecture in the course of reading acquisition. CDP demonstrated that the distinction between phonological assembly and phonological lexical retrieval can be realised in the form of connectivity (either direct of mediated) between orthographic input and phonological output patterns. Thus, CDP had a dual-route architecture but it maintained the uniform computational style of parallel distributed processing (PDP) models, thereby dispensing with any explicit rule-based processing system.

Zorzi et al. (1998b) and Houghton and Zorzi (2003) discuss the dual-route connectionist architecture in relation to the standard multilayer network that is typically employed by connectionist modellers (e.g., Plaut et al., 1996; Seidenberg & McClelland, 1989). Multilayer networks are a generalisation of the feedforward perceptron of Rosenblatt (1962; see Rumelhart et al., 1986, for discussion). Multilayer networks have greater representational power than two-layer networks in which the input and output domains are directly connected (Figure 1A). The use of an intermediate layer of hidden units lying between input and output permits the learning (in principle) of arbitrary nonlinear mappings. A typical multilayer network is built from a two-layer network not only by the addition of hidden units and the necessary connections, but also by the removal of the existing direct connections between the input and output layers (Figure 1B). However, if the hidden units are added but the direct connections are not removed, the network will still be multilayer, but with two distinct pathways from input to output, one direct and the other mediated by hidden units (Figure 1C).

This latter type of architecture has been largely neglected by connectionist modellers. Although it was presented in the seminal work of Rumelhart et al.

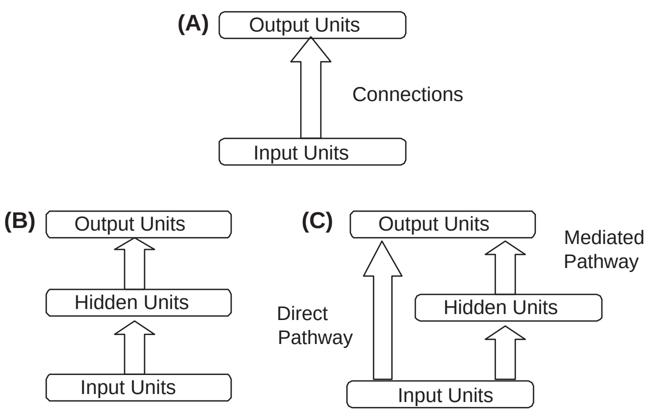

Figure 1. (A) Simplest feedforward network, consisting of two layers of units and connections from input to output units. (B) Single-route multilayer architecture: Hidden units are added between the input and output units and the direct connections from input to output are removed. (C) Dual-route multilayer architecture: Hidden units are added but the direct input–output connections are not removed.

(1986) as the simplest solution to the exclusive-or (XOR) problem, all subsequent simulations were based on networks in which direct connections were not allowed. The multilayer dual-route architecture has properties that distinguish it from the more common single-route version of multilayer networks. First, learning takes place in both pathways at the same time, but the network tends to partition the learning such that the direct pathway will learn simple (linear) regularities, while the mediated route will respond to idiosyncratic (exception) input–output pairs by recognising the exceptional inputs and correcting the regular response produced by the direct pathway (Zorzi et al., 1998b). In this case, the network's ability to generalise to novel stimuli tends to be concentrated in the direct pathway. Second, damage to the two pathways has different effects, so that double dissociations between regular items and exceptional items (i.e., regular words and nonwords vs. exception words) can be observed (Zorzi et al., 1998b). The production of such dissociations has proved extremely challenging for connectionist models based on the standard multilayer architecture (Bullinaria & Chater, 1995; Plaut et al., 1996).

Plaut and colleagues (1996) defined as "quasiregular" any domain in which a limited number of idiosyncratic exemplars (i.e., exceptions) coexist with a large number of "regular" exemplars that can be described (at least in principle) by a set of rules. The decomposition of the task between regular and exception in the architecture of Figure 1C provides a strong computational motivation for connectionist dual-route approaches to learning in quasiregular domains. The roots of this idea can be found in Rumelhart et al.'s (1986) XOR example, where a single hidden unit acted as a feature detecting the conjunction of the two inputs to switch off the output unit (i.e., the exception to the linear OR mapping). Indeed, the CDP approach is not limited to reading aloud (Perry et al., 2007; Zorzi et al., 1998b), but it has been successfully extended to spelling (Houghton & Zorzi, 2003) and to English inflectional morphology (Thomas & Karmiloff-Smith, 2002). Importantly, the dissociation between regular and exception in a CDP architecture does not imply subscription to the traditional (Chomskian) view that productivity in language is based on abstract symbolic rules (e.g., Clahsen, 1999; Marcus, Brinkmann, Clahsen, Wiese, & Pinker, 1995). The dissociation is rather an emergent property of the network architecture in response to the statistical properties of the training set (also see Perry, Ziegler, Braun, & Zorzi, this issue 2010, for further discussion of the rule vs. statistics debate). The apparent "rule-governed" behaviour showed by the direct pathway in a CDP architecture is entirely driven by the existence of reliable input–output mappings and it would not develop if the mapping were arbitrary, such as in the case of the mapping between print and meaning.

THE CDP MODEL OF READING ALOUD

Zorzi, Houghton, and Butterworth (1998a, 1998b) studied in great detail the performance of a simple two-layer associative network (i.e., without hidden units) trained on a set of about 3000 monosyllabic words to learn the mapping between orthography and phonology. Zorzi et al. found that this network acquires properties that are considered the hallmark of a phonological assembly process—they therefore named it the Two-Layer Assembly (TLA) network. Learning in the TLA network does not require the back-propagation algorithm but only the simpler delta rule learning procedure (Widrow & Hoff, 1960). The delta rule is equivalent to a classical conditioning law (the Rescorla-Wagner rule; see Sutton & Barto, 1981, for a formal demonstration) and it has been employed by a number of authors to account for human learning (see Siegel & Allan, 1996, for review). 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. However, the positions are defined with respect to orthographic onset (i.e., letters preceding the vowel letter) and orthographic rime (or word body, i.e., all letters from the vowel onwards). The phonological representation has a similar format, with phonemes in a syllable aligned to phonological onset and rime positions.

Zorzi et al. (1998b) showed that the TLA network is able to extract the statistically more reliable spelling–sound relationships in English, without forming representations of the individual training items (such as the exception words). Therefore, the phonological assembly route in the CDP model produces regularised pronunciations (if the input word is an exception word) and is not sensitive to the base frequency of the trained words. The model provides a good match to the nonword reading performance of human subjects, and can also read single letters and graphemes. The direct connections between letters and phonemes in the TLA network can sometimes be "read out" as pronunciation rules (e.g., initial letter M is always pronounced as /m/), but most of the sublexical spelling–sound mappings discovered by the network are sensitive to the local context and their size is variable (see Zorzi et al., 1998b). Indeed, the output of the network reflects the relative consistency of a given mapping. In agreement with the fact that the major locus of inconsistency in pronouncing English words is the vowel (e.g., EA in HEAD, MEAL, GREAT; e.g., Treiman, Mullennix, Bijeljac-Babic, & Richmond-Welty, 1995), the TLA network, along with the most common mapping of the vowel, delivers other alternative, less common mappings, which are activated to a lesser extent.

The emergent behaviour of the model corresponds to proposals which had been made before. For instance, Brown and Besner (1987) suggested that some individual letters or graphemes may be associated with a small number of phonological realisations. The possibility of rules with multiple outputs has also been proposed by Patterson and Morton (1985) in relation to the "body subsystem" contained in their model. Finally, the idea of multiple outputs is incorporated in Carr and Pollatsek's (1985) notion of "islands of reliability" in spelling–sound correspondence, which implies that correspondences are more reliable for some graphemes than for others. Thus, the system would typically specify a single output for consonants and multiple outputs for vowels. All these proposals are consistent with the behaviour of the TLA network, but they contrast sharply with the model of phonological assembly endorsed by Coltheart et al. (2001) in the DRC model, where the Grapheme-to-Phoneme Conversion (GPC) route is a production system based on a set of explicit rules specifying the dominant (e.g., most frequent) relationships between letters and sounds.

As already noted, the TLA network has only direct connections between input and output and is therefore incapable of learning inconsistent (or even idiosyncratic) mappings like those contained in exception words. Thus, the correct pronunciation of exception words requires a second pathway with greater computational and representational power. This can be achieved by adding an intermediate (hidden) layer between input and output representations. Word-specific information can be represented over hidden units either in a distributed way (when the two pathways are part of a single network, as discussed in the previous section) or in a localist fashion by allocating each unit to a single word (i.e., word nodes). Thus, the lexical pathway in the model can be conceptualised as an interactive activation network (Coltheart et al., 2001; see Houghton & Zorzi, 2003, for an alternative implementation in the context of the lexical pathway for spelling), or alternatively, by any network that develops mediated, internal representations for the known words. The advantage of a (localist) interactive activation model of the lexical route is that visual word recognition (e.g., perceptual identification and lexical decision task) can be readily simulated (Grainger & Jacobs, 1996). Clearly, with regard to word recognition and retrieval of lexical phonology, any advantage of distributed over localist representations has yet to be demonstrated (see Page, 2000, for an extensive discussion).

In the final CDP model, depicted in Figure 2, the lexical route is not fully implemented. When the input string matches one entry in the model's lexicon, the corresponding phonological word form is directly activated and lexical phonological activation spreads to the phoneme output nodes. This activation (excitatory for phonemes that make up the word and inhibitory for other phonemes) is weighted by the log-scaled frequency of the word and is pooled with the assembled phonology produced by the TLA network. Thus, the phoneme output system is the point of interaction between the two routes, where the model's final pronunciation is produced. The structure of

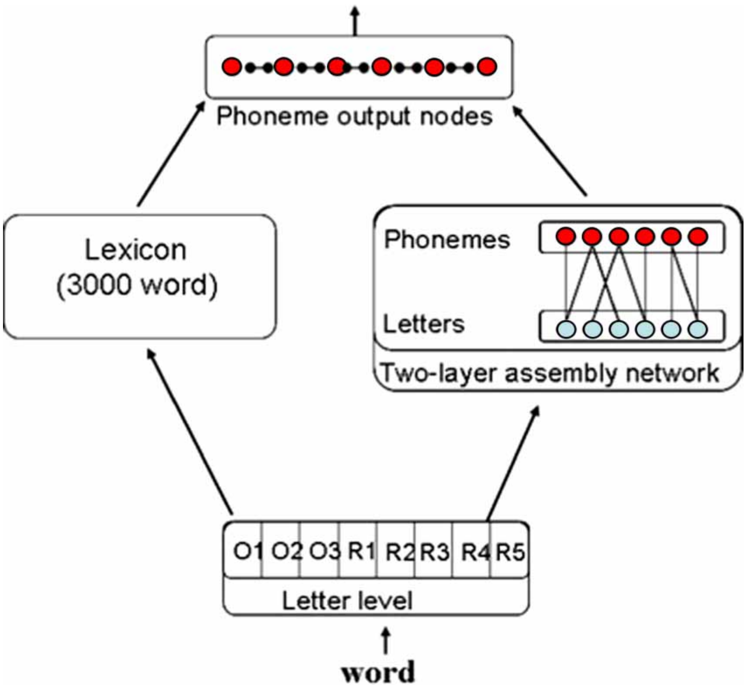

Figure 2. Architecture of CDP (adapted from Zorzi et al., 1998b). The letter level is organised according to a slot-based template, with letters assigned to three onset (O) and five rime (R) positions. Phonemes have a similar alignment into onset and rime slots, both in the TLA network and in the phoneme output buffer.

this system is identical to the output layer of the TLA, but incorporates features (lateral inhibition and gradual activation decay) that provide a temporal dynamic and sensitivity to response competition caused by alternative mappings, which is postulated to be a causal factor in naming latencies. Activations change over time until one of the units in each activated phoneme group reaches a response threshold. The time that the network takes to settle is taken as a measure of naming latency.

The CDP model has been shown to account for the main empirical phenomena in reading aloud, such as the effects of frequency, lexicality, consistency, and regularity, etc. (see Coltheart et al., 2001, and Zorzi, 2005, for comparative evaluations of the model). CDP has also been shown to provide a good match to the performance of patients with surface dyslexia as a result of simulated damage to its lexical route (Zorzi et al., 1998b). Finally, the acquisition of spelling–sound mappings in the sublexical route of CDP during learning provides an excellent account of developmental data on reading acquisition (Hutzler, Ziegler, Perry, Wimmer, & Zorzi, 2004; Zorzi et al., 1998a).

NESTED INCREMENTAL MODELLING: FROM CDP TO CDP+

The development of the successor of CDP was motivated by the attempt to build on its strengths and address its shortcomings. This *nested incremental modelling* strategy advocated in Perry et al. (2007) dictates that a new model should be related to or include at least its own direct predecessors and that it should also be tested against the data sets that motivated the construction of the old models before it is tested against new data sets (see Grainger & Jacobs, 1996, for a discussion of this strategy in the context of modelling visual word recognition).

In particular, CDP showed a length effect for nonwords that was too weak in comparison to the human data (Weekes, 1997) and the correlation between model RTs and human RTs on nonwords of varying length was rather poor (Coltheart et al., 2001). The second weakness, shared by all other major computational models, was the modest item-level correlation between model RTs and human RTs on large-scale databases (Coltheart et al., 2001). Spieler and Balota (1997), who collected naming latencies of 2870 words, suggested that successful models should pass two critical tests: First, the amount of variance predicted by computational models should be at least as strong as the strongest correlating single factor. Second, that the amount of variance predicted by computational models should be similar to the correlation derived from factors that are typically shown to be involved in reading, such as log word frequency, orthographic neighbourhood, and orthographic length. These factors accounted for 21.7% of the variance of word naming latencies in the human data. Unfortunately, CDP accounted for 7.73% of the variance of the human naming latencies, DRC for 3.49%, and the Triangle model for 2.54% (Coltheart et al., 2001). Similar figures were obtained on the Wayne State database (Treiman et al., 1995), which contains RTs for all monosyllabic words that have a consonant-vowel-consonant phonological structure.

The new model, CDP+ (Figure 3), was built according to the nested incremental modelling strategy discussed earlier. First of all, CDP needed to be augmented with a fully implemented lexical route. The choice was to implement a localist lexical route that was as close as possible to that of DRC and based on the interactive activation model of McClelland and Rumelhart (1981). The advantage of this solution is that it inherits the strengths of the interactive activation model in capturing effects related to orthographic processing and lexical access, as previously shown in the simulation studies of Coltheart et al. (2001) and Grainger and Jacobs (1996). Note that CDP+ contains a much larger lexicon than CDP. Indeed, the lexicon basically consists of all monosyllabic words, with 7383 unique orthographic patterns and 6663 unique phonological patterns (extracted

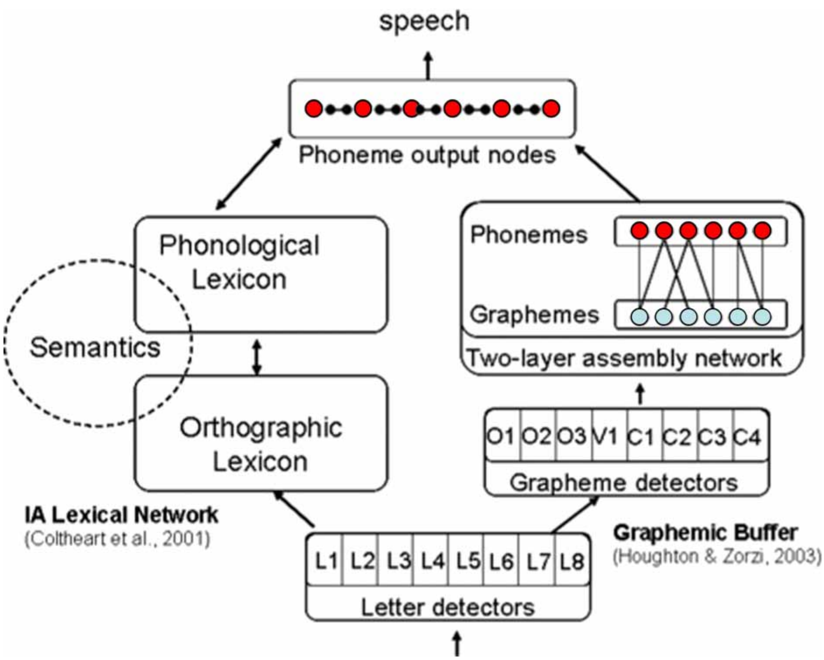

Figure 3. Architecture of CDP+ (adapted from Perry et al., 2007). Letters (L) are encoded in their absolute spatial positions, whereas graphemes are organised according to a graphosyllabic template, with three grapheme slots for the onset (O), one for the vowel (V), and four for the coda (C) positions. Phonemes have a similar alignment into onset, vowel, and coda slots, both in the TLA network and in the phoneme output buffer.

from the English CELEX word form database; Baayen, Piepenbrock, & van Rijn, 1993). The upgraded TLA sublexical network was trained on this larger database.

A further problem for CDP was the relatively high error rate in reading difficult nonwords. A hard test of nonword reading (i.e., generalisation) for connectionist models is the "whammy" set (Rastle & Coltheart, 1998), which includes many nonwords that are highly dissimilar from trained words (e.g., FOOPH, GOICH, TAWSH). CDP had an error rate of around 50% on this set. One way of improving nonword reading in connectionist models is to reduce the "dispersion" (Plaut et al., 1996) of spelling–sound relationships across slot positions. The use of better input and output representations implies that the frequency at which the same letters map onto the same phonemes is generally increased, which in turns facilitates learning of the most common statistical relationships. In CDP+, this was achieved by using graphemes as orthographic input for the TLA network instead of single letters. The level of grapheme representation was added to the model by implemented the graphemic buffer of Houghton and Zorzi's (2003) connectionist dual route model of spelling. The primary motivation for the assumption that the representation is structured into "graphosyllables", with onset, vowel, and coda constituents (Caramazza & Miceli, 1990; Houghton & Zorzi, 2003) comes from the studies of patients with a specific acquired disorder of the graphemic buffer (e.g., Caramazza, Miceli, Villa, & Romani, 1987; Cotelli, Abutalebi, Zorzi, & Cappa, 2003). Also, the data from normal readers suggest that graphemes are functional units above the letter level (Martensen, Maris, & Dijkstra, 2003; Rey & Schiller, 2005; Rey, Ziegler, & Jacobs, 2000). The use of a graphosyllabic template as input coding scheme boosted nonword reading accuracy. The error rate on the "whammy" nonwords dropped from 50% in CDP to only 2.1% in CDP+.

Grapheme units are the input level of the TLA network, but graphemes must be first computed on the basis of the information that is available at the letter level in CDP+ (which, in turn, is activated by letter features) and then inserted into the appropriate slot of the graphemic buffer. Accordingly, a graphemic parsing process that is controlled by focused spatial attention segments letters into graphemes to be submitted to the TLA network. Complex graphemes (e.g., TH) are selected over simple graphemes (e.g., T + H) and graphemes are fully activated when inserted in the graphemic buffer, even if their constituent letters are not when taken from the letter level. The latter implies that CDP+ can account for additive effects of stimulus degradation and length or word frequency (Besner & Roberts, 2003; O'Malley & Besner, 2008; see Ziegler, Perry, & Zorzi, 2009, for CDP+ simulations), because the sublexical part operates as if thresholded processing were used at the letter level.

To illustrate the grapheme parsing mechanism, take for example the word check. Based on a syllabic representation, the first largest grapheme encountered, CH, should be assigned to the first onset position of the graphemic buffer. The next grapheme among the remaining letters (ECK) is the vowel letter E, which should be assigned to the vowel position. The remaining two letters, CK, correspond to a single grapheme that should be assigned to the first coda position. Whether phonological assembly is best conceived as a serial or parallel process has been a hotly disputed issue. Indeed, data taken to support the DRC model's assumption of serial letter processing in the GPC route (e.g., Coltheart & Rastle, 1994; Rastle & Coltheart, 1999) could be accounted for by the parallel CDP model (Zorzi, 2000; but see Roberts, Rastle, Coltheart, & Besner, 2003, for a subsequent study in which CDP did not account for the serial effect). It is important to emphasise that serial processing in CDP+ is not instrumental to accounting for serial effects but it is motivated by the problem of graphemic parsing, a processing stage that was not envisaged in CDP. The timing of phonological assembly in CDP+ is influenced by the serial grapheme parsing process (i.e., the first grapheme is available to the TLA network before the second, and so on) but phonological assembly remains intrinsically parallel because activation spreads between graphemes and phonemes regardless of how many graphemes are available at a given time step. Nonetheless, serialising the sublexical route was found to be important for simulating the length effect on nonwords in CDP+ and it did also significantly improve the correlation with word naming latencies on large-scale databases of real word reading. In a broader theoretical context, the serial processing assumption in CDP+ is specifically linked to spatial attention mechanisms that control attention shifts from left to right over the letter string. This hypothesis is supported by studies showing that experimental manipulations of spatial attention influence phonological assembly but not (or much less) lexical processing (e.g., Reynolds & Besner, 2006; Sieroff & Posner, 1988). In the same vein, deficits of spatial attention after brain damage (i.e., hemispatial neglect) affect phonological assembly but not lexical processing (e.g., Ladavas, Shallice, & Zanella, 1997; Sieroff, Pollatsek, & Posner, 1988). Impaired orienting of spatial attention is strongly correlated with defective phonological decoding skills in children with developmental dyslexia (Facoetti et al., 2006, 2010).

To summarise, CDP+ combines the sublexical network of CDP (updated with graphemic representations and serial graphemic parsing) with the localist lexical network of DRC. As in CDP, the point of interaction between the two routes is the phonological output buffer, a competitive network where lexical and sublexical phonological codes are pooled online to drive the final pronunciation (compare Figures 2 and 3). Phonological representations are identical in CDP and CDP+, with phoneme output nodes syllabically organised into onset-vowel-coda positions.

Perry et al. (2007) proposed a list of 13 empirical phenomena in word naming that constitute the benchmark effects for computational models of reading aloud (Table 1), partly based on a previous list proposed by Coltheart et al. (2001). CDP+ accounts for all these effects, which also include two important forms of acquired dyslexia (see Denes, Cipolotti, & Zorzi, 1999, for a review). Moreover, it accounts for about 20% of the variance in the large-scale databases of human naming latencies. In particular, CDP+ accounts for as much variance as the three factors mentioned in Spieler and Balota (1997; orthographic length, frequency, orthographic neighbourhood). In sum, CDP+ is greatly superior to both its predecessor CDP and its competitors as a model of reading aloud. CDP+ has not been used yet to account for lexical decision data. However, given that the lexical routes in CDP+ and DRC are identical from the feature level up to and including the orthographic lexicon, it should be fairly easy to replicate the simulations of lexical decision performed by

TABLE 1
Benchmark effects for computational models of reading aloud (from Perry et al., 2007)

Name of effect	*Description*
Frequency	High-frequency words are faster/more accurate than low-frequency words
Lexicality	Words are faster/more accurate than pseudowords
Frequency × Regularity	Irregular words are slower/less accurate than regular words. Jared (2002) reports no interaction with frequency
Word consistency	Inconsistent words are slower/less accurate than consistent words. The size of the effect depends on friend/enemy ratio
Nonword consistency	Nonword pronunciations show graded consistency effects, that is, people do not always use the most common grapheme–phoneme correspondences
Length × Lexicality	Naming latencies increase linearly with each additional letter
Position of irregularity	The size of the regularity effect is bigger for words with first position irregularities (e.g., chef) than for words with second- or third-position irregularities
Body neighbourhood	Words with many body neighbours are faster/more accurate than words with few body neighbours
Masked priming	Words preceded by an onset prime are faster/more accurate than words preceded by unrelated primes
Pseudohomophone advantage	Nonwords that sound like real words (e.g., bloo) are faster/more accurate than orthographic controls
Surface dyslexia	Specific impairment of irregular word reading, which is modulated by the consistency ratio of the words
Phonological dyslexia	Specific impairment of nonword reading, which is reduced when nonwords are orthographically similar pseudohomophones
Large-scale databases	Naming latencies of the model are regressed onto the average naming latency of each item in large-scale databases containing thousands of items

Coltheart et al. (2001). This is clearly an advantage of the nested incremental modelling strategy.

FFCDP+: A SIMPLE FEEDFORWARD VERSION OF CDP+

In one of the analyses designed to investigate which parts of CDP+ were responsible for its improved performance (Perry et al., 2007, Part V), CDP+ was turned into a feedforward model by eliminating the recurrent, top-down connections as well as the activation of orthographic neighbours. Strikingly, the feedforward version of CDP+ accounted for an almost identical proportion of item-level variance to that accounted for by the normal version of CDP+. This result is extremely interesting for two main theoretical reasons. First, it shows that the fully implemented lexical route (inherited from DRC) does not add much over and above the effect of a frequency-weighted activation of lexical phonology. Second, it is a strong argument against the criticism that the superior performance of CDP+ is simply due to the large number of free parameters (e.g., Sibley, Kello, & Seidenberg, this issue 2010).

Perry et al. (2007) also reported that the feedforward version of CDP+ (henceforth ffCDP+) showed the same sensitivity to spelling–sound consistency of the full model. This is not surprising because consistency effects related to orthographic units of different grain sizes (e.g., graphemes, word bodies) have been shown to emerge in the TLA network (Zorzi, 2000; Zorzi et al., 1998a,b). However, ffCDP+ was not tested and compared to CDP+ in a more systematic way. Therefore, besides the issue of item-level variance accounted for, a more comprehensive investigation of ffCDP+ is interesting in its own right. Following the recent proposal of Yap and Balota (2009), the analysis presented here is based on regressing the model (and human) latencies onto key word recognition variables (frequency, length, etc.), rather than simply looking at quantitative fits between model and human data. This analysis can reveal whether the model's latencies are influenced to the same extent by the variables that affect the human latencies and provides a much more fine-grained test for the model's quantitative fit than the percentage of variance accounted for. For these analyses I employed the database of 2428 monosyllabic words used by Balota, Cortese, Sergent-Marshall, Spieler, and Yap (2004). The corresponding human naming latencies were obtained from the English Lexicon Project (ELP; Balota et al., 2007) because item means computed over the large number of ELP participants (rather than over 30 participants in the original study by Balota et al., 2004) imply a smaller error variance (and potentially an increased amount of reproducible variance; see Rey, Courrieu, Schmidt-Weigand, & Jacobs, 2009). Matching this database against the CDP+ lexicon resulted in a set of 2385 monosyllabic words with human latencies that could be used for the simulations (from now on referred to as the Balota dataset).

In the simulations reported later, I adopted the same parameter set used in Perry et al. (2007) to turn CDP+ into a feedforward model. First, ffCDP+ has no recurrent connections. This was achieved by setting all feedback parameters to zero (for both lateral and top-down connections). Second, the activation of orthographic neighbours was prevented by setting a strong inhibition between the letter level and the orthographic lexicon (the parameter was increased from -0.55 to -1.0). This ensures that no orthographic neighbours of a word are ever activated in the lexical route. Third, the strength of the excitatory connections between phonological lexicon and phonological output buffer was slightly increased (from .128 to .135) to compensate for a slower build-up of activation in the phonological output buffer that results from the absence of recurrent processing.

Note that the lexical route of ffCDP+ has 7 parameters, whereas the parameter-heavy lexical route of CDP+ has 15 parameters. It could be argued that the number of parameters in ffCDP+ is still large. However, the parameters are not free but they are set in a way that the lexical route simply activates the correct lexical phonology. Moreover, the entire lexical route, from the letter level up to and including the phonological lexicon, could be replaced by a frequency-weighted lexical phonological activation with only two parameters, that is, excitation and inhibition of phoneme nodes in the output buffer (exactly as in the older CDP model; Zorzi et al., 1998b). To provide a more formal demonstration of this claim, I turned off the sublexical route in ffCDP+ (by setting to zero the activation sent by the TLA to the output buffer) and submitted the Balota et al. (2007) dataset to the model. All words were named correctly and, crucially, the model latencies were entirely driven by word frequency. The negative correlation between model latencies and scaled log frequency values (used in the phonological lexicon of the model) was near-perfect, $r = -.98$, $p < .0001$.

Model latencies for this item set were then collected from both ffCDP+ and CDP+. After the removal of phonological errors and RT outliers (items that yielded RTs longer than three standard deviations from the mean), the final data set contained 2352 words that had latencies for ffCDP+, CDP+, and human participants. This ensures that all subsequent analyses were performed on identical sets of words. The correlation between ffCDP+ and CDP+ latencies was extremely high, $r = .995$, $p < .001$, $N = 2352$, suggesting that the performance of the two models is very similar despite the different processing dynamics of the respective lexical routes. This is corroborated by the fact that the correlation between model latencies and ELP human latencies was virtually identical for the two models, $r = .403$ for ffCDP++ and $r = .409$ for CDP+. Note that the correlations between ffCDP+ latencies and the original data of Balota et al. (2004), which included latencies for both young and older participants, were .39 and .419, respectively.

The naming latencies (ffCDP+, CDP+, and ELP human data) were then submitted to regression analyses that used as predictors the four most important lexical variables (cf. Balota et al., 2004):

- *Word frequency*. The logarithm of the HAL frequency norms (Lund & Burgess, 1996).
- *Word length*. The number of letters in the word.
- *Neighbourhood size*. The number of orthographic neighbours that can be obtained by changing one letter (i.e., Coltheart's N).
- *Body consistency*. A consistency measure computed for the word body using Jared's (2002) equation based on the frequency of friends and enemies.

The human data were analysed using a hierarchical regression, whereby 13 descriptors of the onset phoneme (dummy-coded articulatory features; see Balota et al., 2004) were entered in the first step and lexical variables were entered in the second step as predictors. The variance associated with voice key biases are therefore removed in the first step, so that the standardised coefficients (beta weights) for the lexical variables obtained in the second step are more directly comparable to those obtained in the analysis of the models' latencies (articulatory effects are outside the scope of most computational models of reading aloud).

The results are presented in Table 2 using standardised regression coefficients (betas) for each predictor variable. Comparing the effects between models, one can conclude that ffCDP+ shows the same sensitivity to the main lexical variables, even in terms of the effect size, as the full-blown

TABLE 2
Standardised beta-coefficients for standard predictor variables from regression analyses on human and model data (items from Balota et al., 2004)

	Standardised regression coefficients (beta)		
Lexical variables	*ffCDP+*	*CDP+*	*Human (ELP)*
Log frequency	−.721**	−.691**	−.272**
Letter length	.275**	.314**	.138**
Orthographic *N*	−.043*	−.048*	−.074**
Rime consistency	−.144*	−.149*	−.106**
R^2	.699	.695	.491†

**p <.001, *p <.01, †including onset coding.

CDP+. The comparison between models and human data shows that all effects go in the correct direction but that the size of the frequency effect is too strong in both models. Therefore, this type of analysis reveals that a different set of parameters might improve the fit between model and human data (see Adelman & Brown, 2008, for a different method for diagnosing the sources of models' mispredictions).

For the purpose of the present study, the important result is that the main lexical effects in word naming can be accounted for by a greatly simplified version of CDP+. The interactive, recurrent, parameter-heavy lexical route does not seem to be necessary to account for the main factors that influence written word naming. These results do not imply that a lexical route is not needed, but only that word-specific (i.e., lexical) information could take any other form. They also do not imply that feedback processing is not necessary. Not all phenomena in word naming can be explained by a purely feedforward model; in particular, the simulation of the pseudohomophone advantage in nonword naming relies on the existence of feedback connections between phoneme output nodes and phonological lexicon (Perry et al., 2007). Feedback is also important to account for context effects in written word perception (McClelland & Rumelhart, 1981). The main theoretical implication of the results with ffCDP+ is that the CDP+ model is intrinsically modular and that the two-layer associative network of phonological assembly is its only core component. Moreover, the constraints on the type of lexical route appear to be minimal, so that the current lexical route could be replaced by a very different type of network. This issue is taken up in the Future Directions section.

Interestingly, Diependaele, Ziegler, and Grainger (this issue 2010) used components of CDP+ to augment the Bimodal Interactive Activation Model (BIAM) of visual word recognition (Grainger & Ferrand, 1994) so that it could perform the reading aloud task. They implemented the sublexical route (graphemic buffer and TLA network) and the phonological output buffer of CDP+ as point of interaction between lexical and sublexical phonology. The augmented BIAM model can account for fast-acting phonological influences during word recognition (masked phonological priming) and at the same time it can accurately read irregular words aloud, a pattern that poses a serious challenge to the DRC model (Rastle & Brysbaert, 2006). Diependaele et al. point out the striking similarity between their model and CDP+, although the BIAM architecture was motivated by very different constraints (i.e., bimodal interactivity in a model of word comprehension). Note that the components they took from CDP+ are those that form the core of ffCDP+.

BEYOND CDP+

There are many different ways in which CDP+ can be further improved and extended. The ongoing work is focused on two main objectives. The first is to develop versions of CDP+ that work with languages other than English. A German version of CDP+ is already available (Perry et al., this issue 2010) and we are developing Italian and French versions. In addition to providing a useful tool to further promote reading research in these other European languages, a systematic comparison between models trained on different languages should provide insights into the issue of whether the features of a specific language can shape the functional (and neural) architecture of the reading system (e.g., Paulesu et al., 2000; Ziegler et al., 2010). For instance, consistency of the mapping from print to sound has an immediate impact on the contribution of the direct input–output pathway (i.e., the TLA network) versus the lexically mediated pathway to the computation of phonology. Strength of lexical processing might therefore be tuned to the orthographic transparency of the language that the model is trained on. Note that most changes required to adapt CDP+ to other languages involve the TLA network, because, in addition to the problem of learning the spelling–sound mappings for a given language, the alignment of graphemes in the input layer must be structured in a way that reflects the statistical properties of the specific orthography. In contrast, adapting the lexical route is simply a matter of replacing the English lexicon with the words contained in a language-specific database (e.g., the German version of CELEX; see Perry et al., this issue 2010).

The second main objective is to develop a multisyllabic version of CDP+. Indeed, at least by type counts, most words in English have more than one syllable (e.g., Baayen et al., 1993). Despite this, CDP+ (as all other influential computational models of reading aloud) has been developed for monosyllabic words. Dealing with multisyllabic words is a tough modelling challenge. Consider the words canon and canal: The model not only needs to know where to put the syllable boundary (ca.non, ca.nal), but also that canon is stressed on the first syllable, whereas canal is stressed on the second syllable. Even assuming that it is possible to look up this information in the phonological lexicon, one still faces the problem that people can read nonwords, such as commoke or zortess (see Rastle & Coltheart, 2000), for which they consistently assign stress on the first syllable in zortess and on the second in commoke. This means that in the absence of lexical phonology, people are able to assign stress nonlexically. Any new model of disyllabic processing should be able to predict such reading patterns.

CDP++, the successor of CDP+, is a dual process model of reading aloud mono- and disyllabic English words with a lexicon of about 32,000 words (Perry et al., in press). Again, most of the changes required to extend

CDP+ to disyllabic words involve the TLA network. In other words, the modelling challenge was to develop a new sublexical route that could learn the spelling–sound mappings for both mono- and disyllabic words. In CDP+, graphemes and phonemes are assigned to specific slots in a left–right manner based on whether they belong to the onset, the vowel, or the coda. In CDP++, the graphosyllabic and phonological templates are duplicated, so that two syllables can be represented. However, a disyllabic word like canal could potentially be segmented as ca.nal or can.al. In the first case, the grapheme -n would go in the onset of the second syllable, whereas in the second case it would go in the coda of the first syllable. A number of different linguistic constraints can be used to segment disyllabic words (Hall, 2006). However, most of these constraints do not have to be explicitly represented. To assign graphemes without lexical information, the graphemic parser in CDP++ uses onset maximisation as its core principle (this is a well-known constraint in phonology). Thus, consonant graphemes occurring between two vowels are assigned to the onset positions of the second syllable, whenever possible. With the word canal, the model would maximise the -n, and hence use the syllabification ca.nal. A second principle used in graphemic parsing comes from internal network dynamics (i.e., what has been learnt in the TLA network). The idea is that the statistical information captured by the network during training provides implicit constraints to the operations of the graphemic parser. In particular, this prevents the parser from inserting graphemes into slots where nothing has been learnt (note that this information is readily available in terms of strength of the connection weights).

Word stress is modelled with CDP++ simply by adding two extra nodes to the output layer of the sublexical network, which represent predictions by the sublexical network about whether stress should fall on the first or second syllable. Thus, the model not only learns relationships between graphemes and phonemes, but also between graphemes and the stress nodes. The sublexical stress nodes send activation to two stress output nodes, placed at the level of the phonological output buffer, which also receive information about lexical stress from the phonological lexicon. Stress output nodes pool lexical and sublexical information in the same way as phoneme output nodes.

As required by the nested incremental modelling approach, CDP++ captures all empirical phenomena that were accounted for by its predecessor CDP+. Moreover, CDP++ can account for a number of effects specific to disyllabic words (e.g., syllable number, consistency of second syllable, stress typicality; see Yap & Balota, 2009) and nonwords (e.g., stress assignment; e.g., Rastle & Coltheart, 2000). Finally, CDP++ accounts for over 36% of the variance on a set of about 18,000 English words from the ELP database (Balota et al., 2007).

FUTURE DIRECTIONS

The development of the CDP model over the last decade shows very clearly that a nested incremental modelling strategy—commonly used in other areas of science but often neglected in psychology—results in better and more powerful computational models. Although the development of CDP represents one of the most successful and comprehensive computational modelling enterprises in cognitive psychology, there are several aspects of the model that still need substantial improvement.

First, there is no learning at all in the lexical route. Using the interactive activation model as the basis for the lexical route, all recent versions of CDP have inherited its merits as well as its shortcomings. One major problem in adding learning to the current lexical route is that word frequency in the interactive-activation framework is modelled by a node-specific threshold value that directly represents the frequency of the item. However, it is possible to conceive frequency as a dynamic effect in the context of a competitive recurrent network where nodes also have self-feedback connections. In Houghton and Zorzi's (2003) model of spelling, each (localist) word node in the orthographic lexicon had an excitatory feedback loop onto itself, giving it the ability to support its own activation. The strength of this feedback, which depends on a parameter (the feedback weight), can be allowed to vary as a function of word frequency. The modulation of a unit's feedback weight could be achieved as part of a competitive learning algorithm: the feedback loop would be strengthened each time a node is activated. Emphasis on word frequency as a dynamic effect is also a key aspect of the SOLAR model of visual word recognition (Davis, 1999). Frequency effects in SOLAR reflect a node bias mechanism, where bias strength for each word node is constantly revised by the learning algorithm as a function of input statistics.

One more radical solution would be to replace the current lexical route with a completely different network that is not a variant of the interactive activation model. The simulations with ffCDP+ discussed here show very clearly that the constraints on the type of lexical route are minimal. At least with regard to the naming task, the constraints are just two: (1) a frequency-weighted spread of activation to the phonological output nodes; and (2) feedback between phonological output nodes and word-level representation of phonology. Several alternative models of visual word recognition have been proposed in recent years, with special emphasis on the issue of letter position coding and lexical orthographic representations (e.g., Davis, 1999, this issue 2010; Gomez, Ratcliff, & Perea, 2008; Grainger & van Heuven, 2003; Whitney, 2001). Moreover, orthographic coding might be substantially different for lexical access and phonological assembly, with a coarse-grained code for fast access to lexical-semantic representations and a fine-grained code for accessing phonology (e.g., Grainger & Holcomb, 2009). Therefore, in the spirit of the nested incremental modelling strategy, the best one among these models could replace the lexical route in CDP+, or at least its "front end".

In the context of learning lexical representations, a further issue that deserves investigation is how these representations interact with semantics. The DRC model (Coltheart et al., 2001) assumes that nodes in the phonological lexicon can be activated both from the orthographic lexicon (direct lexical route) or via word meanings (lexical-semantic route). The slower lexical-semantic processing would explain why a semantic variable like imageability influences naming only in the case of the slowest items (i.e., low-frequency exception words; Shibahara, Zorzi, Hill, Wydell, & Butterworth, 2003; Strain, Patterson, & Seidenberg, 1995). A different possibility is that orthography makes contact to a level of representation that is neither phonological nor semantic, but intermediate between the two. This view is endorsed in the "Junction" model of Kello (2006), where orthography maps onto a set of hidden units that bind together phonological and semantic representations (note, however, that the Junction model uses the same route to perform phonological assembly, thus departing from the idea that there are two separate ways to get to phonology from print). The idea of a common representation linking phonology, semantics, and orthography is broadly consistent with the lemma level in models of speech production (Levelt, Roelofs, & Meyer, 1999; also see Zorzi, Perry, Ziegler, & Colheart, 1999, on learning lexical-semantic representations).

Finally, another way in which learning in CDP should be improved is by greatly reducing the amount of supervised learning. All connectionist models of reading to date have learned the task of reading aloud through the exposure to a very large corpus of spelling–sound pairs. That is, the input (spelling) and the "desired" output (target pronunciation) for many thousands of words are typically presented until the error-correction procedure employed as learning algorithm reaches a level of performance that is considered adequate by some external criterion. This training regimen is highly implausible: The kind of supervised learning used in all models implies that a teacher externally supplies the pronunciation of all words that should be learnt. The idea that several thousands of words can be taught by externally supplying the correct pronunciation is flawed for a great number of reasons (see Share, 1995, for discussion). Therefore, we need to develop models that are constrained to learn in realistic stimulus environments using a learning regimen that has a sound psychological basis. For example, Dufau et al. (this issue 2010) used unsupervised learning and a realistic corpus (reading materials used in French primary schools to teach kids to read) to map prelexical orthographic representations onto whole-word orthographic representations in a self-organising map (Kohonen, 1982).

Moreover, the simulation study of Hutzler et al. (2004) provides a very effective demonstration that statistical learning by itself, without considering how reading is taught at school, is not sufficient to successfully account for the developmental data on learning to read. Thus, success in this endeavour would open the possibility of assessing the impact of different teaching methods, both for normal children and in remedial treatment of reading disorders.

REFERENCES

Adelman, J. S., & Brown, G. D. A. (2008). Methods of testing and diagnosing model error: Dual and single route cascaded models of reading aloud. *Journal of Memory and Language*, *59*, 524–544.

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

Balota, D. A., Cortese, M. J., Sergent-Marshall, D. S., Spieler, D. H., & Yap, M. J. (2004). Visual word recognition of single-syllable words. *Journal of Experimental Psychology: General*, *133*, 283–316.

Balota, D. A., Yap, M. J., Cortese, M. J., Hutchison, K. A., Kessler, B., Loftis, B., et al. (2007). The English Lexicon Project. *Behavioral Research Methods*, *39*, 445–459.

Besner, D., & Roberts, M. A. (2003). Reading nonwords aloud: Results requiring change in the dual route cascaded model. *Psychonomic Bulletin and Review*, *10*, 398–404.

Brown, P., & Besner, D. (1987). The assembly of phonology in oral reading: A new model. In M. Coltheart (Ed.), *Attention and performance XII* (pp. 471–489). Hove, UK: Lawrence Erlbaum Associates Ltd.

Bullinaria, J. A., & Chater, N. (1995). Connectionist modelling: implications for cognitive neuropsychology. *Language and Cognitive Processes*, *10*, 227–264.

Caramazza, A., & Miceli, G. (1990). The structure of graphemic representations. *Cognition*, *37*, 243–297.

Caramazza, A., Miceli, G., Villa, G., & Romani, C. (1987). The role of the graphemic buffer in spelling: Evidence from a case of acquired dysgraphia. *Cognition*, *26*, 59–85.

Carr, T. H., & Pollatsek, A. (1985). Recognizing printed words: A look at current models. In D. Besner, T. G. Waller, & G. E. MacKinnon (Eds.), *Reading research: Advances in theory and practice* (Vol. 5 pp. 1–82). San Diego, CA: Academic Press.

Clahsen, H. (1999). Lexical entries and rules of language: A multidisciplinary study of German inflection. *Behavioral and Brain Sciences*, *22*, 991–1060.

Coltheart, M. (1978). Lexical access in simple reading tasks. In G. Underwood (Ed.), *Strategies of information processing* (pp. 151–216). London: Academic Press.

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., & Rastle, K. (1994). Serial processing in reading aloud: Evidence for dual-route models of reading. *Journal of Experimental Psychology: Human Perception and Performance*, *20*, 1197–1211.

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.

Cotelli, M., Abutalebi, J., Zorzi, M., & Cappa, S. F. (2003). Vowels in the buffer: A case study of acquired dysgraphia with selective vowel substitutions. *Cognitive Neuropsychology*, *20*, 99–114.

Davis, C. J. (1999). *The Self-Organising Lexical Acquisition and Recognition (SOLAR) model of visual word recognition*. Unpublished doctoral dissertation, University of New South Wales, Australia.

Davis, C. J. (2010). SOLAR versus SERIOL revisited. *European Journal of Cognitive Psychology, 22*, 695–724.

Denes, F., Cipolotti, L., & Zorzi, M. (1999). Acquired dyslexias and dysgraphias. In G. Denes & L. Pizzamiglio (Eds.), *Handbook of clinical and experimental neuropsychology* (pp. 289–317). Hove, UK: Psychology Press.

Diependale, K., Ziegler, J. C., & Grainger, J. (2010). Fast phonology and the Bimodal Interactive Activation Model. *European Journal of Cognitive Psychology, 22*, 764–778.

Dufau, S., Lété, B., Touzet, C., Glotin, H., Ziegler, J. C., & Grainger, J. (2010). A developmental perspective on visual word recognition: New evidence and a self-organising model. *European Journal of Cognitive Psychology, 22*, 669–694.

Facoetti, A., Trussardi, A. N., Ruffino, M., Lorusso, M. L., Cattaneo, C., Galli, R., et al. (2010). Multi-sensory spatial attention deficits are predictive of phonological decoding skills in developmental dyslexia. *Journal of Cognitive Neuroscience, 22*, 1011–1025.

Facoetti, A., Zorzi, M., Cestnick, L., Lorusso, M. L., Molteni, M., Paganoni, P., et al. (2006). The relationship between visuo-spatial attention and nonword reading in developmental dyslexia. *Cognitive Neuropsychology, 23*, 841–855.

Gomez, P., Ratcliff, R., & Perea, M. (2008). The overlap model: A model of letter position coding. *Psychological Review, 115*, 577–600.

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., & Holcomb, P. J. (2009). Watching the word go by: On the time-course of component processes in visual word recognition. *Language and Linguistics Compass, 3*, 128–156.

Grainger, J., & Jacobs, A.M. (1996). Orthographic processing in visual word recognition: A multiple read-out model. *Psychological Review, 103*, 518–565.

Grainger, J., & van Heuven, W. J. B. (2003). Modeling letter position coding in printed word perception. In P. Bonin (Ed.), *The mental lexicon* (pp. 1–24). New York: Nova Science.

Hall, T. A. (2006). English syllabification as the interaction of markedness constraints. *Studia Linguistica, 60*, 1–33.

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

Harm, W. M., & Seidenberg, M. S. (2004). Computing the meanings of words in reading: Cooperative division of labor between visual and phonological processes. *Psychological Review, 111*, 662–720.

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

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*, 273–296.

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

Kello, C. T. (2006). Considering the junction model of lexical processing. In S. Andrews (Ed.), *From inkmarks to ideas: Current issues in lexical processing*. Sydney: Psychology Press.

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

Ladavas, E., Shallice, T., & Zanella, M. T. (1997). Preserved semantic access in neglect dyslexia. *Neuropsychologia, 35*, 257–270.

Levelt, W. J. M., Roelofs, A., & Meyer, A. S. (1999). A theory of lexical access in speech production. *Behavioral and Brain Sciences, 22*, 1–38.

Lund, K., & Burgess, C. (1996). Producing high-dimensional semantic spaces from lexical co-occurrence. *Behavior Research Methods, Instruments, and Computers*, *28*, 203–208.

Marcus, G. F., Brinkmann, U., Clahsen, H., Wiese, R., & Pinker, S. (1995). German inflection: The exception that proves the rule. *Cognitive Psychology*, *29*, 189–256.

Martensen, H., Maris, E., & Dijkstra, T. (2003). Phonological ambiguity and context sensitivity: On sublexical clustering in visual word recognition. *Journal of Memory and Language*, *49*, 375–395.

McClelland, J. L., & Rumelhart, D. E. (1981). An interactive activation model of context effects in letter perception: Part 1. An account of basic findings. *Psychological Review*, *88*, 375–407.

O'Malley, S., & Besner, D. (2008). Reading aloud: Qualitative differences in the relation between stimulus quality and word frequency as a function of context. *Journal of Experimental Psychology: Learning Memory, and Cognition*, *34*, 1400–1411.

Page, M. (2000). Connectionist modelling in psychology: A localist manifesto. *Behavioural and Brain Sciences*, *23*, 443–467.

Patterson, K. E., & Morton, J. (1985). From orthography to phonology: An attempt at an old interpretation. In K. E. Patterson, J. C. Marshall, & M. Coltheart (Eds.), *Surface dyslexia: Neuropsychological and cognitive studies of phonological reading* (pp. 335–359). Hove, UK: Lawrence Erlbaum Associates Ltd.

Paulesu, E., McCrory, E., Fazio, F., Menoncello, L., Brunswick, N., Cappa, S., et al. (2000). A cultural effect on brain function. *Nature Neuroscience*, *3*, 91–96.

Perry, C., Ziegler, J. C., Braun, M., & Zorzi, M. (2010). Rules versus statistics in reading aloud: New evidence on an old debate. *European Journal of Cognitive Psychology*, *22*, 798–812.

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.

Perry, C., Ziegler, J. C., & Zorzi, M. (in press). Beyond single syllables: Large-scale modeling of reading aloud with the Connectionist Dual Process (CDP++) model. *Cognitive Psychology.*

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

Rastle, K., & Brysbaert, M. (2006). Masked phonological priming effects in English: Are they real? Do they matter? *Cognitive Psychology*, *53*, 97–145.

Rastle, K., & Coltheart, M. (1998). Whammies and double whammies: The effect of length on nonword reading. *Psychonomic Bulletin and Review*, *5*, 277–282.

Rastle, K., & Coltheart, M. (1999). Serial and strategic effects in reading aloud. *Journal of Experimental Psychology: Human Perception and Performance*, *25*, 482–503.

Rastle, K., & Coltheart, M. (2000). Lexical and nonlexical print-to-sound translation of disyllabic words and nonwords. *Journal of Memory and Language*, *42*, 342–364.

Rey, A., Courrieu, P., Schmidt-Weigand, F., & Jacobs, A. M. (2009). Item performance in visual word recognition. *Psychonomic Bulletin and Review*, *16*, 600–608.

Rey, A., & Schiller, N. O. (2005). Graphemic complexity and multiple print-to-sound associations in visual word recognition. *Memory and Cognition*, *33*, 76–85.

Rey, A., Ziegler, J. C., & Jacobs, A. M. (2000). Graphemes are perceptual reading units. *Cognition*, *75*, B1–B12.

Reynolds, M., & Besner, D. (2006). Reading aloud is not automatic: Processing capacity is needed to generate a phonological code. *Journal of Experimental Psychology: Human Perception and Performance*, *32*, 1303–1323.

Roberts, M. A., Rastle, K., Coltheart, M., & Besner, D. (2003). When parallel processing in visual word recognition is not enough: New evidence from naming. *Psychonomic Bulletin and Review*, *10*, 405–414.

Rosenblatt, F. (1962). *Principles of neurodynamics.* New York: Spartan.

Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1986). Learning internal representations by error propagation. In D. E. Rumelhart & J. L. McClelland (Eds.), *Parallel distributed processing: Explorations in the microstructure of cognition: Vol. 1. Foundations* (pp. 318–362). Cambridge, MA: MIT Press.

Rumelhart, D. E., & McClelland, J. L. (1982). An interactive activation model of context effects in letter perception: Part 2. The contextual enhancement effects and some tests and extensions of the model. *Psychological Review*, *89*, 60–94.

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

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

Shibahara, N., Zorzi, M., Hill, M., Wydell, T., & Butterworth, B. (2003). Semantic effects in word naming: Evidence from English and Japanese Kanji. *Quarterly Journal of Experimental Psychology*, *56A*, 263–286.

Sibley, D. E., Kello, C. T., & Seidenberg, M. S. (2010). A connectionist model of monosyllabic and bisyllabic naming. *European Journal of Cognitive Psychology*, *22*, 650–668.

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

Sieroff, E., Pollatsek, A., & Posner, M. (1988). Recognition of visual letter strings following injury to the posterior visual spatial attention system. *Cognitive Neuropsychology*, *5*, 427–449.

Sieroff, E., & Posner, M. (1988). Cueing spatial attention during processing of words and letters strings in normals. *Cognitive Neuropsychology*, *5*, 451–472.

Spieler, D. H., & Balota, D. A. (1997). Bringing computational models of word naming down to the item level. *Psychological Science*, *8*, 411–416.

Strain, E., Patterson, K., & Seidenberg, M. S. (1995). Semantic effects in single-word naming. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *21*, 1140–1154.

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

Thomas, M. S. C., & Karmiloff-Smith, A. (2002). Are developmental disorders like cases of adult brain damage? Implications from connectionist modelling. *Behavioral and Brain Sciences*, *25*, 727–788.

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.

Weekes, B. S. (1997). Differential effects of number of letters on word and nonword naming latency. *Quarterly Journal of Experimental Psychology*, *50A*, 439–456.

Whitney, C. (2001). How the brain encodes the order of letters in a printed word: The SERIOL model and selective literature review. *Psychonomic Bulletin and Review*, *8*, 221–243.

Widrow, G., & Hoff, M. E. (1960). Adaptive switching circuits. In *Western electronic show and convention record* (Pt. 4, pp. 96–104). Institute of Radio Engineers.

Yap, M. J., & Balota, D. A. (2009). Visual word recognition of multisyllabic words. *Journal of Memory and Language*, *60*, 502–529.

Ziegler, J. C., Bertrand, D., Tóth, D., Csépe, V., Reis, A., Faísca, L., et al. (2010). Orthographic depth and its impact on universal predictors of reading: A cross-language investigation. *Psychological Science*, *21*, 551–559.

Ziegler, J. C., Perry, C., & Zorzi, M. (2009). Additive and interactive effects of stimulus degradation: No challenge for CDP+. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *35*, 306–311.

Zorzi, M. (1999). *The connectionist dual-process model: Development, skilled performance, and breakdowns of processing in oral reading*. Unpublished doctoral thesis, University of Trieste, Italy.

Zorzi, M. (2000). Serial processing in reading aloud: No challenge for a parallel model. *Journal of Experimental Psychology: Human Perception and Performance*, *26*, 847–856.

Zorzi, M. (2005). Computational models of reading. In G. Houghton (Ed.), *Connectionist models in cognitive psychology* (pp. 403–444). Hove, UK: Psychology Press.

Zorzi, M., Houghton, G., & Butterworth, B. (1998a). The development of spelling–sound relationships in a model of phonological reading. *Language and Cognitive Processes*, *13*, 337–371.

Zorzi, M., Houghton, G., & Butterworth, B. (1998b). 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., Perry, C., Ziegler, J., & Coltheart, M. (1999). Category-specific deficits in a self-organizing model of the lexical-semantic system. In D. Heinke, G. W. Humphreys, & A. Olson (Eds.), *Connectionist models in cognitive neuroscience* (pp. 137–148). London: Springer-Verlag.