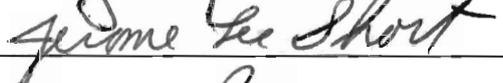
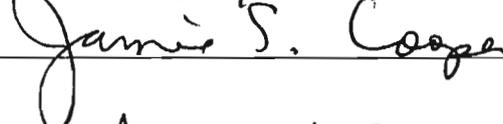


LARGE SCALE MODELING OF SINGLE WORD READING AND RECOGNITION

by

Daragh E. Sibley  
A Dissertation  
Submitted to the  
Graduate Faculty  
of  
George Mason University  
in Partial Fulfillment of  
The Requirements for the Degree  
of  
Doctor of Philosophy  
Psychology

Committee:

	Director
	
	
	Department Chairperson
	Program Director
	Dean, College of Humanities and Social Sciences

Date: August 1, 2008

Summer Semester 2008  
George Mason University  
Fairfax, VA

Large scale modeling of single word reading and recognition

A dissertation submitted in partial fulfillment of the requirements for the degree of  
Doctor of Philosophy at George Mason University

By

Daragh E. Sibley  
Master of Arts  
George Mason University, 2008

Director: Christopher T. Kello, Professor  
Psychology

Summer Semester 2008  
George Mason University  
Fairfax, VA

Copyright: 2008 Daragh E. Sibley  
All Rights Reserved

## DEDICATION

This dissertation is dedicated to my parents, for all the support and opportunities.

## ACKNOWLEDGEMENTS

I would like to thank the faculty of George Mason University's Psychology department, particularly Dr. Christopher Kello, for advice and guidance throughout my graduate education.

## TABLE OF CONTENTS

	Page
List of Tables.....	vi
List of Figures.....	vii
Abstract.....	viii
Chapter 1.....	1
Chapter 2.....	20
Chapter 3.....	35
Chapter 4.....	45
Chapter 5.....	50
Chapter 6.....	63
Appendices.....	72
Work cited.....	81

## LIST OF TABLES

Table	Page
Table 1. Item-variance accounted for by various models on different stimuli sets .....	52
Table 2. Analysis of $R^2$ in naming latencies .....	56
Table 3. Analysis of $R^2$ in lexical decision latencies .....	57

## LIST OF FIGURES

Figure	Page
Figure 1. An orthographic Sequence Encoder .....	8
Figure 2. Schematics of the Triangle, DRC, and CDP+ models .....	11
Figure 3. Orthographic Sequence Encoder performance, by length .....	23
Figure 4. Architecture of Simulation 1 .....	25
Figure 5. Naming accuracy for Simulation 1, as a function of length .....	30
Figure 6. Architecture of Simulation 2 .....	37
Figure 7. Naming accuracy for Simulation 2, as a function of length .....	40
Figure 8. Word and nonword distributions .....	43
Figure 9. Architecture of Simulation 3 .....	46
Figure 10. Naming accuracy for Simulation 3, as a function of length .....	47
Figure 11. The Junction Theory of word reading and recognition .....	66

## ABSTRACT

### LARGE SCALE MODELING OF SINGLE WORD READING AND RECOGNITION

Daragh E. Sibley, PhD

George Mason University, 2008

Dissertation Director: Christopher T. Kello

The study of word reading and recognition has been strongly influenced by computational cognitive modeling. These models facilitate theorizing about the mechanisms that underlie word reading and recognition (e.g., Morton, 1970; McClelland & Rumelhart, 1981; Seidenberg & McClelland, 1989; Plaut, McClelland, Seidenberg, and Patterson, 1996; Harm & Seidenberg, 1999; Coltheart, Rastle, Perry, Langdon, & Ziegler, 2001; Perry, Ziegler, & Zorzi, 2007). However, the preeminent models in this field only process monosyllabic words. This results from difficulties inherent in representing the orthography and phonology of multisyllabic words. To address this issue Sibley, Kello, Plaut, & Elman (in Press) created a connectionist architecture named the Sequence Encoder. The present work utilizes representations from a Sequence Encoder to build models that address an order of magnitude more data than previous models.

A second goal of this work is to explore the possibility of hypothesizing fewer mechanisms in models of the reading system. The three preeminent models of reading all implement two distinct pathways from orthography to phonology. A sublexical route

encodes statistical relationships between letters and phonemes, while a lexical route encodes whole word information. This dissertation explores whether each of these pathways are necessary for word reading and recognition. We present three models trained on 60,000 mono- and multisyllabic English words. Simulation 1 maps from orthography to phonology using a single sublexical route. It demonstrates substantial naming capacities, but is incapable of addressing lexical decision data. Simulation 2 utilizes only a lexical route, where reading is achieved by an inductive process that utilizes whole word information stored in a lexicon. This model addresses naming and lexical decision data on an unprecedented scale. Simulation 3 integrates sublexical and lexical routes from the previous models, but exhibits negligible capacities beyond Simulation 2. Finally, we examine each simulation's sensitivity to stimuli characteristics that impact behavioral latencies. Our simulations mimicked the effects of all examined variables on participants' latencies. These simulations demonstrate that models can be scaled up without incorporating new mechanisms specifically to address phenomena of multisyllabic word reading, such as stress assignment. We conclude that a single lexical pathway from orthography to phonology is sufficient to simulate word reading and recognition.

## CHAPTER 1: INTRODUCTION

Computational models are valuable tools for theory building, particularly for determining which hypothetical cognitive processes best account for an empirical dataset. This is apparent in the lexical processing literature, where progress over the last three decades is marked by the development and refinement of computational models (e.g., Morton, 1970; McClelland & Rumelhart, 1981; Seidenberg & McClelland, 1989; Plaut, McClelland, Seidenberg, & Patterson, 1996; Harm & Seidenberg, 1999; Coltheart, Rastle, Perry, Langdon, & Ziegler, 2001; Perry, Ziegler, & Zorzi, 2007). These models constrain theorizing about the necessary components of the cognitive system, by demonstrating which mechanisms are sufficient to achieving word reading and recognition. More generally, lexical behaviors have become test beds for theorizing about general properties of cognitive systems (McClelland & Rumelhart, 1981; Fodor & Pylyshen, 1988; Pinker and Prince, 1991; Plaut, McClelland, Seidenberg, and Patterson, 1996; Sibley & Kello, 2004).

Three preeminent theories of lexical processing have been instantiated respectively as the Triangle model (Seidenberg & McClelland, 1989; Plaut, McClelland, Seidenberg, and Patterson, 1996; Harm & Seidenberg, 1999), the Dual Route Cascaded (DRC) model (Coltheart et al., 2001), and the Connectionist Dual Process+ model

(CDP+; Perry et al., 2007) of lexical processing. Each model offers a distinct conceptualization and has contributed to our understanding of the lexical system. The Triangle model demonstrated how domain-general principles of parallel distributed processing can account for a broad range of phenomena regarding word reading. The DRC model simulated word reading and recognition within a single framework, offering an accessible conceptualization of reading impairments. The CDP+ model currently predicts the most behavioral phenomena and data, using an almost entirely connectionist architecture.

Despite notable advances, there is a large volume of unexplained behavioral data regarding single word recognition and reading. In particular, the English Lexicon Project (ELP; Balota, Cortese, Hutchison, Neely, Nelson, Simpson, & Treiman, 2002) contains mean data regarding more than 40,000 English words, most of which are multisyllabic. The DRC, Triangle, and CDP+ models process only monosyllabic words, and hence cannot account for the bulk of ELP data. The present work explores simulations of word reading and recognition that are designed to address an order of magnitude more data than previous models. The development of these simulations is guided by a central tenet in the philosophy of science: Occam's razor. This principle emphasizes that if two theories address similar volumes of data, then the more parsimonious theory is preferable. With respect to cognitive modeling, the most parsimonious theory is presumably the one that utilizes the fewest mechanisms and parameters. We focus on Occam's razor to minimize the introduction of additional mechanisms and parameters as models are scaled up.

This paper proceeds by discussing why extant models only address a subset of an adult's vocabulary and how recent advances permit an expansion beyond the current scale. We then review the Triangle, DRC, and CDP+ models and discuss how each model utilizes two distinct processing pathways between orthography and phonology. A sublexical pathway relies on statistical relationships between letters and phonemes, whereas a lexical pathway is sensitive to whole word information. We examine whether both of these pathways are necessary, by examining whether either pathway alone is sufficient to simulating the recognition and reading of 60,000 mono- and multisyllabic words.

Our first model shows that a sublexical pathway alone can support mono- and multisyllabic word naming, but this model is unable to address lexical decision data. A second simulation shows how a model with only a lexical pathway can address both naming and lexical decision data on a large scale. Our third model combines the sublexical and lexical pathways, yet it does not provide a much better account of the data than the model with only a lexical route. We discuss whether hypothesizing both mechanisms is warranted, given that a single lexical route seems sufficient to address a large volume of data. We conclude that the lexical pathway alone may be a viable theory of lexical processing in the context of large-scale modeling. We discuss some avenues for future work to continue investigating this possibility, and to expand the scale and scope of lexical modeling.

### *Scaling up lexical models*

The restriction of current models to monosyllabic word reading results from difficulties inherent in representing the orthography and phonology of multisyllabic wordforms. These difficulties arise from the use of slot based representational schemes, where wordform elements (letters or phonemes) are placed in slots that correspond to wordform positions. As a result, each element is treated independently across slots and so representations cannot capture graded similarities between letters and phonemes occurring in different positions in words of different lengths. Plaut et al. (1996) discussed how this *dispersion problem* makes it difficult to extract regularities that exist within and between a word's written and spoken forms.

Representational schemes have been designed to alleviate the dispersion problem. For instance, CDP+ and the Triangle model, as implemented by Plaut et al. (1996) and Harm & Seidenberg (1999), aligned words to vowel centered templates. This reorganization facilitates the extraction of dependencies between text and speech. However, vowel centering breaks down for longer, multisyllabic wordforms because of increased variability in wordform length. Also, it is unclear which vowels should be used to align multisyllabic words (for further discussion see Sibley, Kello, Plaut, & Elman, 2008).

The restricted scales of current models produce problems when theorizing about the reading and recognition system. The processing of monosyllabic words does not reflect the full complexity inherent in processing of multisyllabic words. For instance, models of monosyllabic word reading do not address how lexical stress is assigned

(Rastle & Coltheart, 2000) or the functional role of the syllable (see Cutler, Mehler, Norris, & Segui, 1986). As a result they cannot address how stress typicality effects word recognition latencies (Yap, 2007), or why syllabic length effects naming latencies beyond the variable of orthographic and phonological length (New, Ferrand, Pallier, and Brysbaert, 2006).

A second problem arises from limiting the behavioral data that are used to evaluate current models. In the absence of large-scale databases, previous models were designed to account for the results of factorial studies (Balota et al., 2004). In these experiments, readers are shown to differentially respond to small groups of words that reflect extremes on some variables of interest. These experiments are often interpreted using level of significance rather than effect size statistics. While these results need to be addressed, models designed exclusively for factorial data may account for relatively little item-variance. Balota et al. (2004) noted that the Triangle model correctly distinguishes between many stimulus sets, but just captures 3.3% of the total item-variance in the benchmark Spieler and Balota (1997) dataset. In fact, the Triangle and DRC models accounted for less item-variance in naming latencies than the logarithmic frequency of each word. The CDP+ model accounts for substantially more item-variance, 17.28% and 21.56%, in the Spieler & Balota (1997) and Balota & Spieler (1998) datasets, respectively. However, CDP+ is still restricted to processing monosyllabic words. In essence, a sensitive to the extremes of a variable does not necessarily produce a sensitive with the correct magnitude relative to other variables.

Balota et al., (2004) offered several reasons why models designed to emulate the results of factorial experiments may have difficulty accounting for variance at the item level. Many of these problems stem from experimental methodologies that take continuous variables like word frequency, and generate groups of stimuli that embody the variables' extremes (Cohen, 1983). Creating extreme groups increases effect sizes in analyses of variance, but it also increases the likelihood of introducing confounding variables (Cutler, 1981). Avoiding confounding variables is particularly challenging in the study of reading behaviors, where many second and third order variable interactions are related to behavior (Yap, 2007).

To alleviate the problem of confounding variables, some studies introduce extensive controlling and counterbalancing stimuli (e.g., Jared, 2002). However, counterbalancing assumes interactions have a linear and monotonic relationship to a behavior and practicalities of experimentation dictate that extensive controlling and counterbalancing limit the number of stimuli in a given experimental condition. As a result, theoretical conclusions are drawn from data regarding even fewer unique stimuli. For instance, utilizing naming latencies for 198 words with 3-6 letters, Weekes (1997) reported no effect of word length after other variables were controlled. However, using a larger corpus, with words that varied between 2 and 8 letters, Balota et al. (2004) observed a strong and unique effect of word length on naming latencies. A final problem with factorial experiments arises from list context effects, where participants' responses may be biased to the extent that they were presented with a biased portion of the stimulus

space. Specifically, Andrews (1997) demonstrated that list context effects can play a substantial role in word recognition studies.

The past several years have seen the development of a large lexical behavioral database, which obviates the development and assessment of computational models using only factorial studies. A joint enterprise of several universities, termed the ELP (Balota et al., 2002) has produced more than two and a half million measurements of single word recognition and more than a million measurements of single word reading. Prior models' abilities to utilize the ELP are limited because most ELP words are multisyllabic. To fully employ the ELP requires models that address multisyllabic words. This necessitated the development of representational schemes that are resilient to the aforementioned dispersion problem.

Recent advances have resulted in the creation of viable representations for multisyllabic words. Sibley, Kello, Plaut, & Elman (in Press) created a connectionist architecture, termed the Sequence Encoder, capable of learning orthographic and phonological representations for nearly 75,000 mono- and multisyllabic English words. These representations are sensitive to the statistical regularities among elements of its corpora. This occurs even when the corpora contain strings that vary substantially in length. As a result, representations created by the Sequence Encoder are resilient to the previously discussed dispersion problem and can be used in lexical models that address mono- and multisyllabic word recognition and reading.

The Sequence Encoder (Sibley et al., in press; Kello, Sibley, & Colombi, 2004) learns to create representations sensitive to the structure in English wordforms under

pressure of an auto-encoding task, i.e. the reproduction of input sequences as output sequences. Auto-encoding drives a connectionist model to discover a means of re-representation that exploits structure in the inputs (Bishop, 1995). For a review of how statistical dependencies affect behavior see Appendix 1. The Sequence Encoder, see Figure 1, is created by concatenating two Simple Recurrent Networks (SRN; Jordan, 1986; Elman, 1990). The first SRN, called the encoding SRN, receives a sequence of inputs and encodes them into a single distributed representation. The second SRN, called the decoding SRN, decodes this representation into a sequence of outputs. The orthographic input representations used for the simulations in this paper are the output activation values which are copied between the encoding and decoding SRNs.

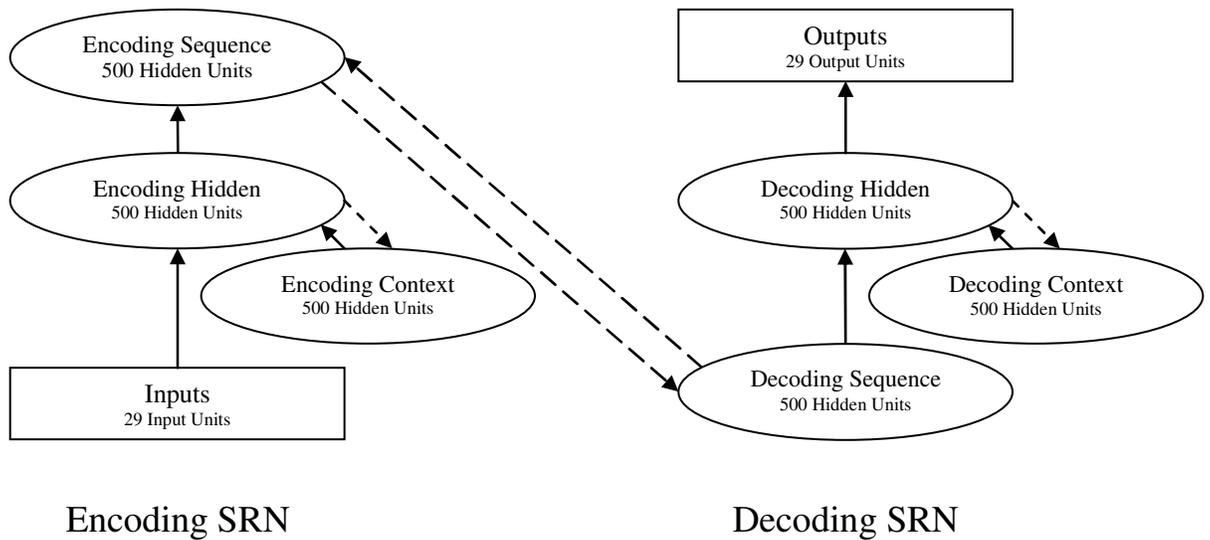


Figure 1. An orthographic Sequence Encoder

Analyses in Sibley et al. (in press) demonstrated that each of the representations learned by a Sequence Encoder is a linear combination of the distributed representations for its constituent letters with respect to their positions in words of different lengths. To illustrate, the representation for CAT# is a linear combination of the representation for: C in the first position of a 3-letter wordform, A in the second position of a 3-letter wordform, T in the third position of a 3-letter wordform, and the end-of-wordform “letter” # occurring at the end of a 3 letter word. Wordform representations created by a Sequence Encoder overcome the dispersion problem because constituent element representations are more similar to the extent that they code similar letters in similar positions, of wordforms of similar lengths.

#### *Extant models of word reading and recognition*

The three simulations presented in this paper use the sequence encoder to build upon prior models of lexical processing. The Triangle, DRC, and CDP+ models all propose that the lexical system is composed of orthographic, phonological, and semantic levels of representation. This is consistent with current neuroscientific evidence. For each theorized representational layers, there is evidence for a cortical structure which performs an analogous operation. The left fusiform gyrus is particularly responsive to visually presented words and pseudowords (McCandliss, Cohen, & Dehaene, 2003). This sensitivity is modal, as spoken words do not produce an analogous metabolic response (Dehaene, Gurvan Clec'h, Poline, LeBihan, & Laurent Cohen, 2002). Also, selective

damage to this region impairs visual, but not auditory word recognition (McCandliss, Cohen, & Dehaene, 2003). This suggests that left fusiform gyrus is participating in the coding of orthographic wordforms representations, akin to the orthographic layer in the present simulations. In contrast, speech elicits bilaterally activation in the superior temporal sulcus (Hickok & Poeppel, 2007), relative to resting baseline and nonspeech sounds. Hemodynamic responses recorded in this region are not affected by whether a stimuli is a word or pseudowords (Binder, Frost, Hammeke, Bellgowan, Springer, Kaufman, & Possing, 2000). However, responses in superior temporal sulcus are modulated by phonological stimuli characteristics like phonological neighborhood density (Okada & Hickok, 2006). This is consistent with the superior temporal sulcus performing a function similar to the phonological wordform layer in the present models. Finally, semantic information is believed to be represented in a highly distributed network throughout neocortex (McClelland, McNaughton, & O'Reilly, 1995).

The Triangle, DRC, and CDP+ models all instantiate two distinct paths between orthographic and phonological levels of representation (See Figure 2). The sublexical route encodes regularities in the mapping between components of orthography and components of phonology, akin to phonics reading instruction. This sublexical route supports the reading of stimuli with regular pronunciations, whether those stimuli are known or novel. The lexical route encodes word-specific information, akin to whole word reading instruction. Consequently this route supports the reading of familiar words, whether they have regular or irregular pronunciations. While the Triangle, DRC, and

CDP+ models all instantiate dual route theories, they offer different interpretations of how processing is achieved in each route.

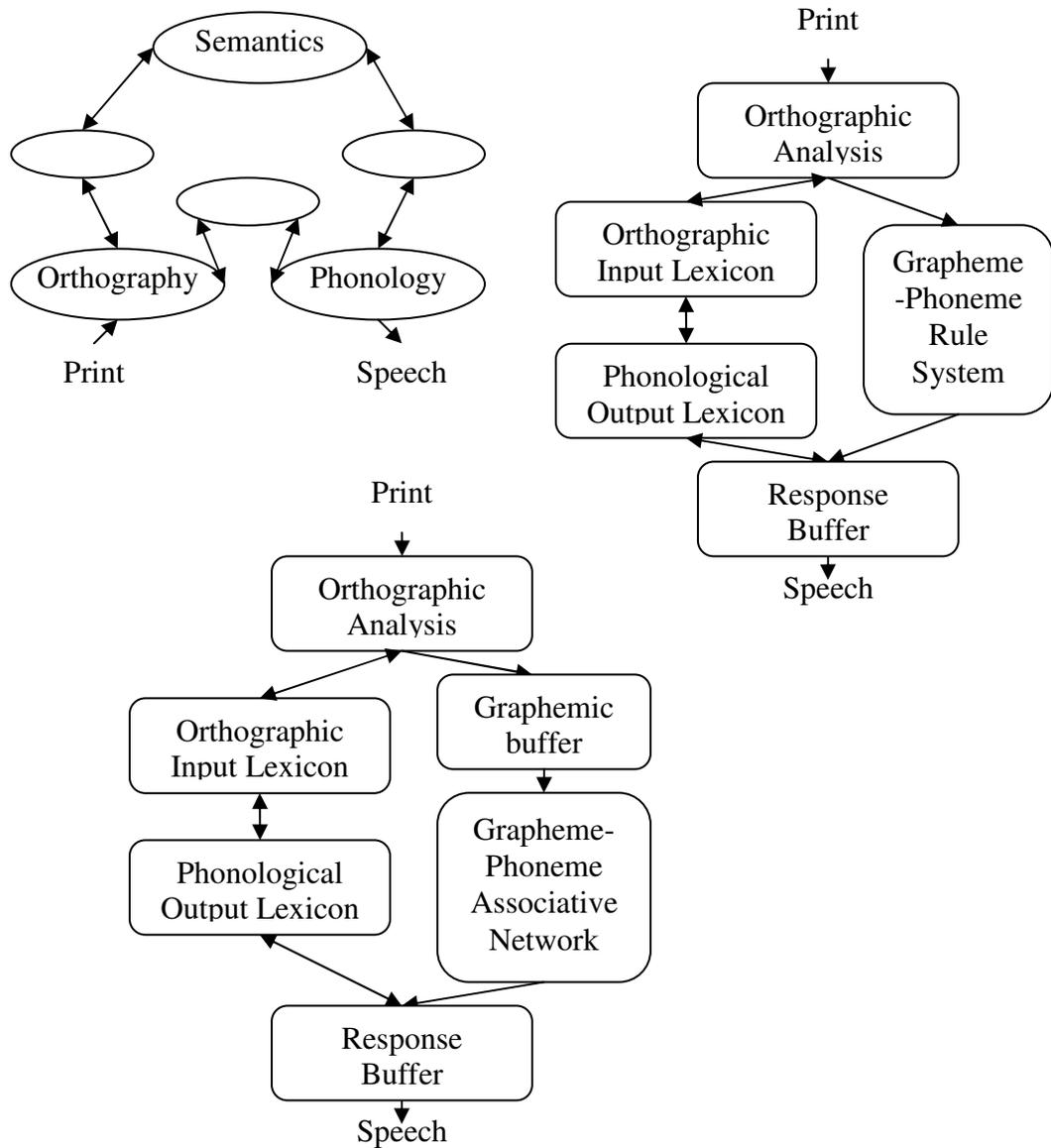


Figure 2: Schematics of the Triangle, DRC, and CDP+ models

The Triangle model of word reading is shown on the top left of Figure 2. Its proponents suggest that reading is achieved by satisfying constraints from orthographic, phonological, and semantic sources (Seidenberg & McClelland, 1989; Plaut et al., 1996; Harm & Seidenberg, 1999). All processing in this model is governed by the same computational principles, those of parallel distributed processing (Rumelhart & McClelland, 1986). However, the two routes of the Triangle model incorporate different constraints. The first route maps directly between orthography and phonology, and so it is sensitive to orthographic and phonological constraints. Because it extracts information about the statistical relationship between particular letters and phonemes, this pathway constitutes a sublexical route. The second route maps between orthography and phonology via a semantic layer of representation. The presence of semantic constraints drives processing to occur with respect to morphemes. Because most monosyllabic words are monomorphemic and this model is only currently applied to monosyllabic words, this route utilizes whole word information and so constitutes a lexical route of pronunciation.

The DRC model (as shown on the top right of Figure 2) proposes that reading is achieved by two cascaded processing routes between print and speech (Coltheart et al., 2001). DRC's sublexical route uses grapheme-to-phoneme correspondence rules to assemble a pronunciation from left to right. DRC's lexical route is an adaptation of McClelland and Rumelhart (1981) interactive activation network, where a letter feature

layer interacts with letter and word layers to identify words that are presented to the model. This orthographic word layer interacts with a second interactive activation network composed of phonological word, phoneme, and phoneme feature levels. The two routes in DRC utilize strictly different computational principles, one tailored to account for expressions of sublexical knowledge (e.g., regularity effects and nonword reading) the other for behavioral expressions of lexical knowledge (e.g., frequency effects and exception word reading). Localist coding is used in the lexical route which enables simulations of lexical decision tasks in addition to word naming tasks.

A diagram representing the CDP+ model (Perry et al., 2007) of word recognition and reading can be found on the bottom of Figure 2. The development of CDP+ followed a nested modeling approach, fully adopting the lexical route of the DRC model. As a result, CDP+ inherits the strengths and weaknesses associated with DRC's lexical route, including identical word recognition performance. CDP+ departs from DRC by replacing the sublexical route, which is a grapheme-to-phoneme conversion algorithm, with a two layer associative connectionist network. Incorporating parallel distributed processing ideologies into both routes offers increased parsimony and avoids some of the limitations of DRC. One important limitation is that DRC's sublexical route has difficulty simulating graded regularity effects, e.g., it can only produce a single pronunciation for a nonword. In contrast, the two layer associative network in CDP+ is more sensitive to the quasi-regular relationship between orthography and phonology.

There are several reasons why extant models implement both sublexical and lexical mechanisms. First, reading education has seen a heated debate about whether

phonics or whole-word instructional techniques are more successful (i.e., Garan, 2002). These educational paradigms directly parallel the distinction between sublexical and lexical reading processes. Second, a number of cognitive and neuroscientific theories suggest that different types of information are stored in different cognitive or biological components. For instance, long term memory research suggests that declarative and procedural information are independently instantiated (Knowlton & Squire, 1993). Pinker's (1999) Words and Rules theory suggests that word information resides in a declarative memory system and explicit rules, instantiated by a procedural system, operate on these word representations. Third, various reading behaviors seem to tap into different types of knowledge. With particular instructions and without time pressure, a proficient reader can demonstrate sublexical knowledge by sounding out a word's pronunciation, or lexical knowledge by distinguishing known from novel stimuli.

While these reasons motivate hypothesizing of separate sublexical and lexical mechanisms, they may not necessitate two separate routes or mechanisms. Ultimately, if single and dual route models account for similar amounts of behavioral data, Occam's razor suggests that single route models are preferable. For this test to be fair, models should simulate behaviors that tap into both sublexical and lexical knowledge. This motivates our efforts to simulate both word reading and recognition. In particular, pseudoword reading is generally believed to require sublexical knowledge, whereas lexical decision is generally believed to require lexical knowledge.

The distinction between sublexical and lexical mechanisms is also expressed in reading deficits. Acquired surface and phonological dyslexias involve a selective deficit

for reading either irregular words or pseudowords, respectively (Marshall & Newcombe, 1973; Behrmann & Bub 1992). The incorrect pronunciations generated by acquired surface dyslexics tend to be regularization errors, in which utterances generated for irregular words conform to the majority pronunciation norms (e.g. PINT is pronounced to rhyme with MINT, LINT, and HINT) (Marshall & Newcombe, 1973). The incorrect pronunciations generated by acquired phonological dyslexics tend to be lexicalization errors, in which pseudowords are pronounced as the most similar known word (e.g. MAVE is pronounced as SAVE; Behrmann & Bub, 1992). The deficit for processing irregular words and impaired lexical decision performance suggests acquired surface dyslexia represents an excessive application of majority pronunciation norms. The deficit for processing nonwords and the production of lexicalization errors suggests that acquired phonological dyslexia represents an excessive application of information specific to particular known words.

Within a dual route framework, acquired dyslexias have been interpreted as strong evidence for separable routes of processing (e.g., Coltheart, Curtis, Atkins, & Haller, 1993; Pinker, 1999). Acquired surface dyslexia is considered the result of damage to the lexical route of reading, while acquired phonological dyslexia is said to result from damage to the sublexical route. In essence, each reading impairment is explained as selective damage to the route that was supporting the impaired cognitive ability.

### *Single-Route Models of Lexical Processing*

The dual route explanation of double dissociations is compellingly transparent, but it is not the only explanation of double dissociations. Shallice (1988) enumerated several ways that double dissociations could occur without isolated damage to separable components of a cognitive system. Patterson & Ralph (1999) argued that acquired dyslexias result from damage to representations rather than routes. Most relevant, Kello and Plaut (2003; Kello, 2003) simulated a double dissociation in a single-route model of word reading.

Kello and Plaut's (2003; Kello, 2003) single route models were motivated by the observation that learning to read scaffolds upon prior learning about speech. Kello and Plaut simulated this process with connectionist models that first learned associations between phonology and semantics. Then the models learned to interface orthographic information with intermediate representations supporting the phonological-semantic associations. This produced a single integrated pathway between orthography, phonology, and semantics. A control parameter, input gain, on connectionist units in the integrated pathway was manipulated to produce the hallmark effects of acquired surface and phonological dyslexia. Input gain is a multiplicative term on a connectionist unit's output activation function that modulates the strength of a unit's response to an input.

Kello, Sibley, & Plaut (2005; Sibley & Kello, 2004) explored how this double dissociation was produced in models computing an abstract quasi-regular function. In Kello et al. (2005) inputs were mapped onto outputs via either distributed or localist intermediate layers of representation. In both types of model, adjusting the input gain on

nodes in the intermediate layers had qualitatively similar effects. Decreasing input gain biased processing to occur with respect to regularities across large numbers of input-output pairs. As a result, irregular items elicited erroneously regular outputs, while regular and novel items were processed correctly. Increasing input gain created an excessive sensitivity to input-output mappings for specific items in the training corpus, which impaired the models ability to process novel items. Kello et al (2005) demonstrates that routes using both distributed and localist representations can be driven between modes that are reminiscent of sublexical and lexical processing.

The localist models used by Kello et al. (2005) were abstracted versions of analogical processing theories that have been proposed to achieve word reading (Glushko, 1979) and explored by Damper & Marchand (2000). Analogy models require individual words to be stored as exemplars. These exemplars delineate known from novel stimuli and so naturally support word recognition. This explicit distinction produces difficulties in explaining how novel stimuli are processed. Nonword naming by analogy requires transferring information from whole source exemplars to target nonwords. To effectively achieve generalization, the model needs sufficient numbers of source exemplars for analogical transfer to all possible well formed nonwords. As a result, naming nonwords by analogy will likely become more plausible as the scale of a model increases.

Kello (2006) presented a large-scale simulation of word reading and recognition inspired by Glushko's (1979) analogy theory. This model processed 45,273 mono and multisyllabic words and to our knowledge was the first attempt at simulating this scale of

behavioral data from the ELP. Orthographic and phonological wordform representations for this model were created by a Sequence Encoder, while semantic representations were created from textual co-occurrence statistics (Rhode, 2007). Wordform representations interacted through a single route of processing that utilized localist coding to create a lexicon, and so constituted a lexical route of processing. Because each lexicon node and wordform representation was pre-specified, a simple wiring algorithm could create connection weights from the lexicon to orthographic, phonological, and semantic layers. Weights were assigned so that each lexicon node became activated when the representations on the orthographic layer was similar to the representation of its corresponding word. In turn, activation of each lexicon node produced the corresponding pre-specified distributed representation on the phonological wordform layers. Lexicon nodes were placed in competition for activation, so that nodes competed to contribute to the processing of each stimulus.

Kello's (2006) model simulated word naming and recognition with only a single lexical route and accounted for far more data than the Triangle, DRC and CDP+ models. This model's chief limitation was that it could not name nonwords, which is the reason to hypothesize a sublexical route. The large scale of this model's lexicon provided numerous exemplar sources, but the hardwired associations between the lexicon and phonological layers did not support analogical transfer. Also, the phonological wordform representations and decoding SRN of the phonological Sequence Encoder did not support analogical transfer in the context of reading. To illustrate, to read the nonword SINT the Kello (2006) model could have utilized source information regarding the words SANE

and FIST. These exemplars each share two letters with the target nonword and together suggest pronunciations for all four letters in SINT. The model from Kello (2006) could not achieve this sophisticated level of analogical transfer. Instead the model would have simply averaged together SANE and FIST's phonological representations, to produce a representation that was unlikely to be decoded into a reasonable pronunciation for SINT. For a connectionist model to be capable of reading nonwords with only a lexical route requires training how to perform analogical transfer without actually training on the novel stimuli.

The present work seeks to scale up computational lexical models while examining whether a single route is sufficient to simulate naming and lexical decision data. To this end, we present three models of single word reading and recognition trained on 60,000 mono- and multisyllabic English words. Simulation 1 maps between orthographic and phonological representations using a single sublexical route designed according to the tenets of parallel distributed processing. Simulation 2 explores the viability of models with only a lexical route, where naming occurs through a process of analogy to words in a lexicon. Finally, Simulation 3 incorporates both sublexical and lexical routes to create a fully connectionist dual route model. These three models permit an examination of the benefits of dual routes over either a lexical or sublexical route alone. After each simulation is presented we will examine their abilities to account for behavioral data regarding large volumes of mono and multisyllabic words.

## CHAPTER 2: SIMULATION 1

Simulation 1 was trained to recognize and read 60,000 mono- and multisyllabic English words. In Simulation 1 a single sublexical route bridged between orthographic and phonological layers of representation. This route is created by a learning algorithm that alters connection weights between unspecified connectionist units to reduce error in outputs. In this respect, Simulation 1 is an adaptation of the direct orthography-to-phonology route from the Triangle model. However, we assess Simulation 1 as a full model of word reading and recognition, rather than as part of a dual route theory. This section focuses on two fundamental questions: Can a model with a single sublexical route read mono- and multisyllabic English words and nonwords at a much larger scale than any extant models? Can such a model also discriminate between mono- and multisyllabic words and nonwords?

### *Simulation 1: Methods*

*Training corpora.* Orthographic and phonological wordforms were selected from the corpus used in Sibley et al. (in press). This corpus was an intersection of the Wall Street Journal corpus (Marcus, Santorini, & Marcinkiewicz, 1993) and the CMU Pronunciation Dictionary (<http://speech.cs.cmu.edu/cgi-bin/cmudict>), which provided phonetic

transcriptions. After removing homographs and words with more than 11 letters, the remaining list was intersected with the ELP database to yield 28,032 words with associated behavioral data.

To demonstrate scalability we included some words from the Sibley et al. (in press) corpus that did not appear in the ELP. Many of these words are aberrant and receive pronunciations that are inconsistent with English pronunciation norms. It should be expected that the Simulation's word naming performance will be artificially low on these items. As a result, naming performance will be reported separately for words that do and do not appear in the ELP. Also, the 28,032 words included in the ELP provide the best test bed of our model's naming performance, as there is information about how often participants correctly read those words.

Learning to read large numbers of words with inconsistent pronunciations may disrupt the reading of nonwords, which should receive pronunciations consistent with majority grapheme-phoneme correspondences. To mitigate this issue the words with the most inconsistent pronunciations were removed. As the majority of these words were multisyllabic, inconsistency was determined by Levenshtein consistency. This metric measures whether words with similar spellings receive similar pronunciations (Yarkoni, Balota, & Yap, submitted). Levenshtein consistency is calculated as the ratio of a word's Levenshtein orthographic distance to its Levenshtein phonological distance. These terms are computed as the average of the minimum number of letters or phonemes that must be added, removed, or substituted to transform a word into its nearest 20 neighbors. Yarkoni et al. (submitted) suggested this metric for corpora with substantial variability in

word lengths. The words with the least consistent pronunciations were discarded, to produce the final training corpus of 60,000 words (Levenshtein consistency is also used for more substantive purposes later on).

*Input and output representations.* Orthographic input representations for Simulation 1 were learned by a Sequence Encoder model trained specifically for the current work. This Sequence Encoder was similar to the model described as Simulation 2 in Sibley et al. (in press). It had two banks of analogous input and output units. One bank consisted of 26 letter units plus one word-delimiting unit and the other bank consisted of two units that identified whether a given letter was a consonant or vowel (“Y” was considered a vowel). A given letter was coded by activating a single unit from each bank.

There were two differences between the new orthographic Sequence Encoder and Simulation 2 in Sibley et al. (in press). First, the sequence element previously representing end-of-wordform was activated at both the beginning and end of each sequence, which meant that letters could appear before or after this sequence element. Second, the new Sequence Encoder reached asymptotic training after 400,000 epochs, while Simulation 2 in Sibley et al. (in press) was trained for 250,000 epochs. The new Sequence Encoder correctly auto-encoded every letter and consonant/vowel unit for 90.17% of the 60,000 words in Simulation 1’s corpus. Figure 3 depicts that errors tended to occur on longer words. The large majority of mistakenly decoded words contained only one incorrect letter (i.e., 85.49%). Thus sequence encoders were successful enough in capturing letters and their positions to be used as inputs to our models of word reading.

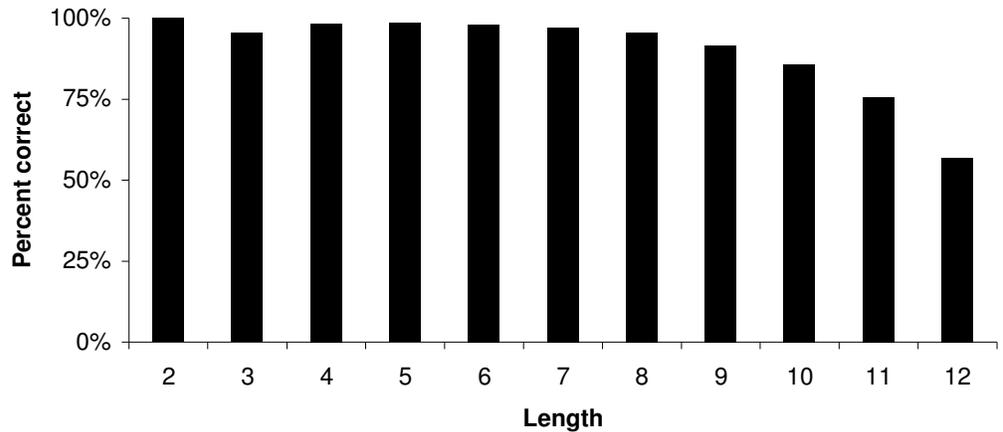


Figure 3. Orthographic Sequence Encoder performance, by length

Instead of creating phonological wordform representations under pressure of an auto-encoding task these representations were learned specifically to support reading. Phonological outputs were created by an output SRN analogous to the decoding SRN of the orthographic Sequence Encoder. This SRN converted a single distributed representation into a series of phonemic outputs coded along two banks of units. The first bank of outputs included 14 vowel units and 25 consonant units plus an end-of-wordform unit. A given phoneme was coded by activating a single unit in each bank. The second output bank consisted of four stress units; three units coded levels of stress with an additional unit coding no stress for consonants.

*Architecture.* As depicted in Figure 4, the sublexical model was constructed by concatenating an SRN onto the output of a feed-forward perceptron. The perceptron transformed a single orthographic wordform representation into a single phonological wordform representation. The SRN decoded each phonological representation into a sequence of phonemic outputs. In Figure 4, solid arrows indicate learned connections between every unit in the sending and receiving layers, with the number of units in each layer displayed within the figure. With one exception, connection weights were initialized by randomly sampling values from a uniform distribution in the range [-0.1, 0.1]. Connection weights emanating from the Hidden Layer were initialized to a reduced range of [-0.025, 0.025] to prevent the large number of connections from creating binary representations very early in training. The dashed arrow in Figure 4 denotes a copy function that SRNs use to buffer previous hidden unit states.

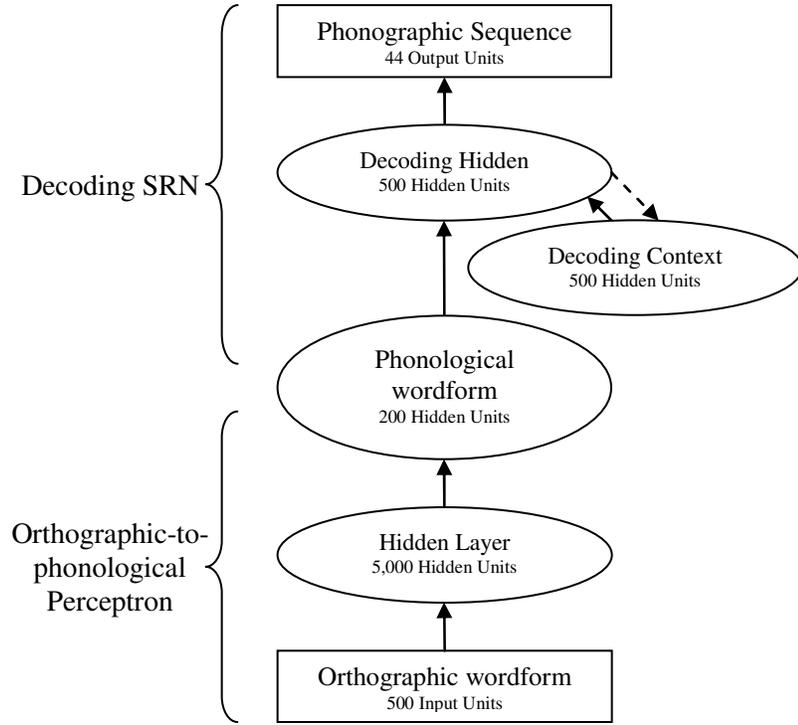


Figure 4. Architecture of Simulation 1

*Model operations.* Each word was processed as a number of discrete time steps. On each time step the decoding SRN produced a different phonemic output, by virtue of the SRN buffering information to alter activations at subsequent time steps. For each layer, the forward propagation of activation began by calculating each unit's net input. Net inputs,  $I_j$ , were calculated as,

$$I_j = \sum_i w_{ij} a_i, \quad (1)$$

where  $w_{ij}$  was the connection weight from sending unit  $i$  to receiving unit  $j$ . Activation values,  $a_j$ , on hidden unit were calculated as,

$$a_j = \tanh(I_j), \quad (2)$$

this hyperbolic tangent function is equivalent to a logistic function, but bounded between [-1, 1]. Activation values on output units were calculated as normalized exponentials,

$$a_j = e^{I_j} / \sum_l e^{I_l}, \quad (3)$$

on each time step output activation was calculated as the exponential of its net inputs, normalized over output units,  $l$ , in each bank.

During training output unit activations were used to generate a training signal, derived from Kullback-Leibler divergence error (Rumelhart, Durbin, Golden, & Chauvin, 1995). The error for each output pattern,  $Error_p$ , was calculated as,

$$Error_p = -\log(a_T) * \sqrt{WSJ_w} / leng_w, \quad (4)$$

where  $a_T$  was the output activation of the target output units for each time step,  $WSJ_w$  was the Wall Street Journal frequency of the word being processed (Marcus et al., 1993), and  $leng_w$  was the words phonemic length. Divergence error can be simplified to the first term in equation 4, because on each time step there was only one node in each bank which should be activated, and it had a target of 1. Divergence error was scaled by the square root of the WSJ frequency (Marcus et al., 1993), to approximate how real language learners have more opportunities to learn about higher frequency words. Finally, each error term was divided by the word's phonemic length to decouple phonemic length and error magnitude. For each time step, error signals were back-

propagated through the decoding SRN and perceptron. Error was attributed to each hidden node,  $\delta_j$ , according to,

$$\delta_j = f'_j(net_j) \sum_{i \in P_j} \delta_i w_{ij}, \quad (5)$$

so that the error on each node was a function of the error on its posterior nodes, the weights connecting these nodes, and the derivative of the node's activation function.

Due to the magnitude and scale of the training corpora, these simulations were too computationally demanding to be feasibly run on conventional personal computing hardware. As a result, we used the Teragrid, a high performance grid computing architecture made accessible by the National Science Foundation. To utilize this resource the model's training regime was designed to operate in parallel across arbitrary numbers of CPUs. Parallelization occurred using a synchronous client-server model, where each training epoch was composed of 4 stages. In stage one, the server CPU broadcasted the current weight matrices for the model and each of the 49 client CPUs loaded them into memory. In stage two; each client CPU processed 50 words, sampled without replacement for each client independently and then error was attributed to each connection weight ( $Error/w_{ij}$ ). In stage three; the server accumulated error information from the clients. Finally, in stage 4; the server utilized error information to change the models' connection weights according to,

$$\Delta w_{ij} = \eta \sum \partial Error / \partial w_{ij}, \quad (6)$$

where  $\eta$  was a learning rate set to  $7.5 \times 10^{-7}$  throughout the model, except for the weights emanating from the Hidden Layer, where  $\eta$  was set to  $7.5 \times 10^{-8}$  to compensate for the

very large number of weights. The reduction in error was judged to asymptote after 10,000 training epochs, at which point training was halted.

Parallelization was necessary because serial training time for the 3 simulations reported in this paper, with current PC hardware would have required roughly 5 years. This estimate does not include any of the models created during iterative model development. The utilized parallelization scheme was effective because the most computationally expensive portion of the back-propagation algorithm is calculating and accumulating error information. Splitting this task across 49 client CPUs reduced the time to perform this stage of processing by a factor of nearly 49. There was very little overhead associated with using more clients, as passing weight matrices between CPUs occurred in trivial time, for our purposes.

#### *Simulation 1: Results and discussion*

Simulation 1's performance was assessed by instantiating an orthographic representation on the model's input layer and propagating activation forward for 14 time steps or until the most active output node was the end-of-wordform unit. For each time step the most active output unit in each bank was taken as the model's output. The resulting phonemic sequence was considered the model's pronunciation for each wordform.

To examine Simulation 1's naming accuracy, the 60,000-item training corpus was divided into two groups. The 28,032 words with corresponding behavioral data in the ELP comprised a Trained ELP group. The remaining 31,968 words without naming data in the ELP constituted the Trained non-ELP group. In Simulation 1 64.5% and 34.3% for

Trained ELP and Trained non-ELP words, respectively, were pronounced identically to the target pronunciation in the CMU Pronunciation Dictionary, including stress and end-of-wordform units. This performance disparity reflects that Trained non-ELP were generally less consistent and also had a lower error scaling term (a word's WSJ frequency divided by its phonemic length). The average of this term was 14.14 for Trained ELP words, but only 3.07 for Trained non-ELP words, so weights were altered more in the service of learning Trained ELP items.

Among the 30,954 words with at least 1 error, 75.28% of the individual phoneme and stress units were correctly produced. Of the 60,000 items, 5,120 were monosyllabic words that were also used by Perry et al. (2007) to assess CDP+. Of these words, 95.97% were named correctly. Generally, phonemic length influenced naming performance, as depicted in Figure 5. Longer words were processed less accurately for a number of reasons. For longer words, error signals had to be backpropogated across more time steps during training and more information compressed into each wordform representation. These are practical difficulties which may be specific to this Simulation. However, two problems are likely to impact any model of reading as it is scaled up to address multisyllabic words. Longer words provide more opportunities to produce incorrect phonemes, and multisyllabic words embody more complex relationships between orthography and phonology relative to monosyllabic words.

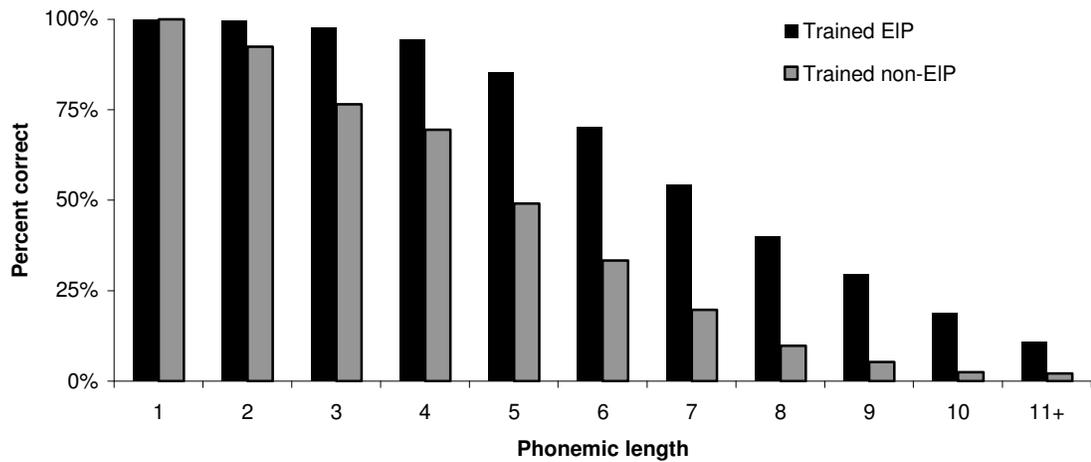


Figure 5. Naming accuracy for Simulation 1, as a function of length

Nonword naming performance was assessed on two different groups of stimuli. The first comprised 552 monosyllabic pseudowords assembled by Seidenberg et al. (1994) that were later used to assess the nonword naming ability of CDP+ in Perry et al. (2007). The second nonword set comprised 600 mono- and multisyllabic nonwords created for Sibley et al. (in press), which consisted of 150 items from each of the four most common orthographic lengths (5-8 letters) in our corpus. These nonwords are provided in Appendix 3.

Nonwords do not have target pronunciations that can be used for assessing Simulation 1's outputs. As a result, we adopted an approach utilized by Seidenberg et al. (1994) and adopted by Perry et al. (2007), where human raters judged whether the pronunciation for each nonword corresponded to grapheme-phoneme relationships existing in any English words. Three independent human raters judged nonword naming

accuracy to be on average 58.64% and 38.11%, for the Seidenberg et al. (1994) and Sibley et al. (in press) nonwords, respectively. Raters also judged the number of nonword pronunciations which could be corrected by a single Levenshtein operation (an insertion, deletion, or substitution of a phoneme). On average raters judged 31.34% and 33.73% of the Seidenberg et al. (1994) and Sibley et al. (in press) nonwords, respectively to be in this class. To examine the internal consistency of raters' judgments, Cronbach's Alpha was calculated across the three raters, where a nonword judged to be correct was coded as a 1, a miss by one operation was coded as 0.5, and an incorrect was coded as a 0. Cronbach's alpha for Simulation 1 was 0.88 for both stimulus groups. George and Mallery (2003) propose a rule of thumb where an alpha above 0.9 is excellent, 0.8 is good, and 0.7 is acceptable.

A viable single route model must simulate behaviors that reflect sublexical and lexical forms of knowledge. In particular, it should simulate behaviors generally hypothesized to require the excluded route. So word recognition abilities, which seemingly tap into whole word knowledge, are an important assay of sublexical models of reading. Simulation 1's sublexical route from orthography to phonology utilizes hidden distributed representations that capitalize on systematic relationships between letters and phonemes. There is evidence that distributed representations may be capable of simulating lexical decision behaviors. In particular, Plaut (1997) showed that the polarity of distributed representations on a semantic output layer could distinguish words from nonwords. Creating a semantic output layer in a single route model requires linking semantics into the distributed representation between orthography and phonology, as in

the Integrated Pathway (Kello, 2003) and Junction models (Kello, 2006). However, semantics were not implemented in Simulation 1, due to the excessive computational demands of learning the unsystematic relationship between semantics and orthography or phonology. While semantics can certainly contribute to word recognition, there is evidence that word recognition may occur without semantic access. So we examined whether distributed patterns of activation on Simulation 1's Hidden Layer could differentiate words from nonwords.

Word recognition abilities are behaviorally assessed with lexical decision tasks. Simulation 1's training corpus contained 27,881 words with accompanying Lexical Decision data in the ELP. The ELP also contains a volume of behavioral data regarding lexical decisions to nonwords, 27,881 of which were randomly selected to match the number of words. As Plaut (1997) demonstrated that polarity of distributed representations can in some situations distinguish words from nonwords, we also used a polarity metric to assess Simulation 1's word recognition abilities. A measure of polarity was created by summing the absolute value of the activation values on the 5000 units of Simulation 1's Hidden Layer (which varied between -1 and 1). Here, words had an average polarity of 2633.85 and standard deviation of 77.64, while nonwords had an average polarity of 2640.75 and standard deviation of 79.55. As a result, it was impossible to place a decision boundary between these two distributions that could effectively distinguish words from nonwords. Given this inability, Simulation 1 was not used to address behavioral data regarding lexical decisions.

Analyses of Simulation 1 reveal the capacity to scale up models of word reading with only a single sublexical route. Simulation 1 correctly named nearly five times as many stimuli as any previous model. Naming accuracy for monosyllabic words was consistent with extant models, and many multisyllabic words were named with acceptable pronunciations. Further, Simulation 1 named many nonwords, some of which were multisyllabic. Yet, the overall percentage of correctly read words and nonwords was well below a proficient reader's abilities. This could result from theoretically uninteresting limitations in this implementation, such as sufficient computational power to learn the mapping from orthography to phonology.

To explore this possibility, a model was trained with the same architecture, but with nearly four times as many connection weights, where connection weights equate roughly to computational power in connectionist networks. This larger network's training asymptoted with a similar level of error and word naming performance, suggesting a fundamental limitation in this architecture or the orthographic wordform representations created by the Sequence Encoder. Orthographic representations were structured exclusively by statistics within orthography, and so may not be sufficient to support word and nonword reading. At the least, there is evidence that children's orthographic representations are influenced by phonological and reading development (Ziegler & Goswami, 2005; Goswami & Ziegler, 2006).

A theoretically important issue arises from Simulation 1's inability to distinguish words from nonwords on the distributed representations of its sublexical route. Simulation 1 is clearly not a fully implemented model of word reading or recognition, as

this must incorporate a semantic layer of representation. In a single route sublexical model, a semantics layer would be linked into the distributed representations bridging between orthography and phonology. Incorporating semantics may produce an ability to distinguish words from nonwords, and so provide a framework for simulating lexical decision behaviors. Connecting semantics to the hidden layer between orthography and phonology would likely engender a serious problem. Semantics constraints would tend to pull apart the distributed representations for words and nonwords on the layer between orthography and phonology. Connectionist models support generalization by similarly representing trained and novel stimuli. As a result, allowing semantics to influence these representations would further compromise this architecture's ability to generalize during word naming. Fundamentally, improving nonword reading performance and simulating lexical decision behaviors seem to be in opposition in a model with a single sublexical route. However, a full test of this assertion requires a model with a fully implemented semantic component.

## CHAPTER 3: SIMULATION 2

DRC and CDP+ achieved word recognition through the use of lexicons that conform to Barlow's (1972) definition of localist coding, where increased activation on a node corresponds to an increasing probability that a particular word was encountered. As a result, these models architecturally differentiate words from nonwords. This creates routes which demonstrate very little sensitivity to sublexical correspondences between orthography and phonology and so do not support behaviors like nonword reading. DRC and CDP+ hypothesize a separate sublexical route (Coltheart et al., 2001). Simulation 2 was designed to examine the necessity of including this sublexical route.

In the discussion of Simulation 1 we suggested that strongly differentiating words from nonwords in the single intermediate representation between orthography and phonology could compromise generalization abilities. Kello (2006) produced a model with a single lexical route that simulated word naming and lexical decision for 45,000 English words, but could not read any nonwords. Nevertheless, Kello et al. (2005) demonstrated that localist representations could in principle support generalization. Unlike the distributed representations used in Simulation 1, localist representations may offer the capacity to distinguish known from novel stimuli while still supporting generalization.

Behavioral evidence suggests pronunciation may occur through a process of analogy to word exemplars (Glushko, 1979). Simulation 2's lexical route was designed to support analogical transfer from known word exemplars to novel strings. This was achieved by allowing for the graded activation of lexicon nodes, so that all stimuli presented to the model elicited a response from numerous lexicon nodes. As the model learned to pronounce these patterns of activation for words, it should learn how information from multiple words could be blended to handle novel stimuli.

#### *Simulation 2: Methods*

Simulation 2 was designed to permit a controlled comparison with the other models presented herein. Simulation 2 used identical corpora and input/output representations to those in Simulation 1.

*Architecture.* As depicted in Figure 6, Simulation 2 was constructed similarly to Simulation 1. The chief difference between the models is that the 5,000 unit Hidden Layer from Simulation 1 was replaced with 60,000 nodes that locally coded for words in the model's training corpus. Each node was instantiated as a radial basis function, whose centroid was the orthographic representation of a different wordform. Effectively, each node acted as a receptive field centered on a particular word. A node became active when an input fell within its receptive field, producing an increased response when the input was closer to its center. Receptive fields allowed multiple nodes to produce graded responses to a single input.

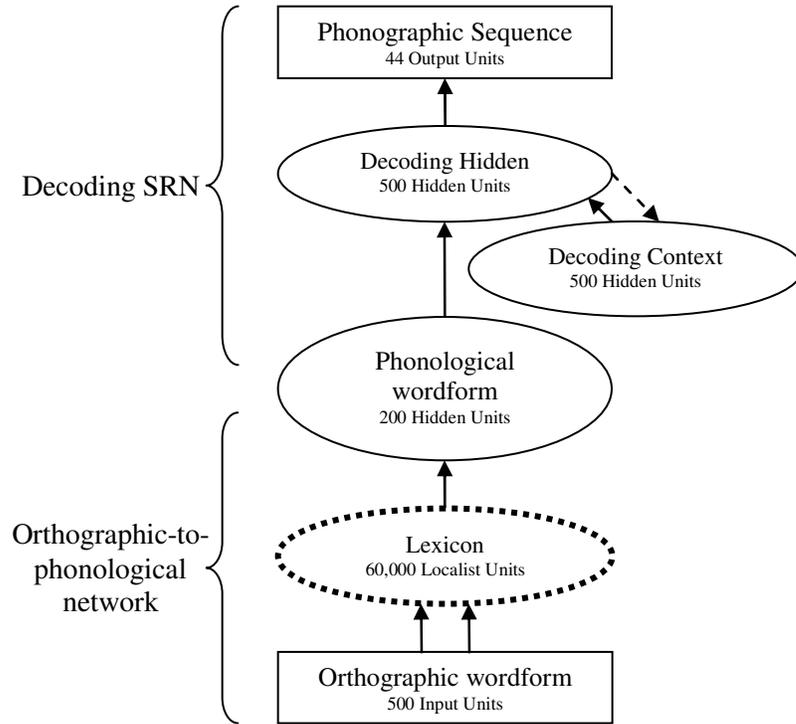


Figure 6. Architecture of Simulation 2

Portions of Simulation 2 that were shared with Simulation 1 were instantiated identically, with the same parameters. The new connection weights emanating from the lexicon were initialized randomly with a range of  $[-0.1, 0.1]$ . The new double arrows in Figure 6 represent connection weights created by a wiring algorithm that followed,

$$w_{ij} = a_{ij}, \quad (7)$$

where  $a_{ij}$  was the activation on the  $i$ th node of the distributed representation of word  $j$ , as determined by the Orthographic Sequence Encoder. Learning algorithms (Grossberg, 1980; Rummelhart, 1995) could also be adapted to this task to create a model of reading acquisition. This wiring algorithm was a useful simplification given that the current models are designed to simulate adult behavior. Each lexicon node's net input was calculated as,

$$I_j = \sum_i (a_i w_{ij})^2, \quad (8)$$

which is the squared Euclidian distance between incoming connection weights and the activation on the Orthographic Wordform layer. The dot product activation function used in other portions of the model was not used here because it generates larger net inputs as a result of more binary activations. For radial basis functions, net inputs should increase as inputs move closer to a centroid. Output activation of each node in the lexicon was calculated as,

$$a_j = Freq_j e^{-\gamma * I_j}, \quad (9)$$

where  $Freq_j$  was the  $\log_{10}$  of the WSJ frequency for word  $j$  and  $\gamma$  is input gain, a multiplicative term that modulated the size of each node's receptive field. Input gain was set to 0.035 which was large enough that most nodes did not become active for every input, but it was small enough that nonwords produced adequate responses to support naming.

*Model operations.* The operations and training of Simulation 2 were identical to Simulation 1, with two exceptions. The learning rate on connection weights emanating from the lexicon was set to 0.00003 and model training was judged to reach asymptote after 5,500 epochs.

#### *Simulation 2: Results and discussion*

Simulation 2's naming performance was assessed identically to Simulation 1's using the same corpora. Performance on the Trained ELP and Trained non-ELP words was 93.7% and 66.2%, respectively. With respect to the 28,032 Trained ELP words, Simulation 2 was nearly as accurate as ELP participants, who averaged 95.38% correct. Of the 5,120 monosyllabic words from Perry et al. (2007), Simulation 2 read 99.7% correctly, which is marginally better than Simulation 1. However, Simulation 2 correctly read 18,383 more words than Simulation 1. Figure 7 illustrates that, as with Simulation 1, phonemic length influenced which items were named correctly. Figure 7 also depicts that performance fell off more slowly than in Simulation 1. Longer words were produced less correctly for the same reasons as in Simulation 1.

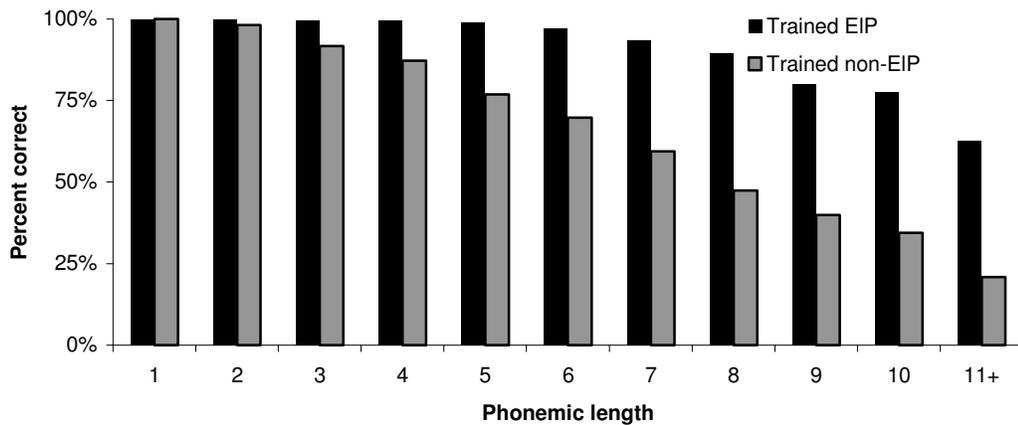


Figure 7. Naming accuracy for Simulation 2, as a function of length

Simulation 2 demonstrates superior word naming performance to Simulation 1, where the only differences between these simulations were that a sublexical layer of representation was replaced with a lexical layer of representation. Because naming performance was relatively similar for monosyllabic words, it suggests that a lexical route created from localist representations may be better suited to supporting the naming of multisyllabic words. This could occur because of qualitative differences in the mapping between orthography and phonology for mono- and multisyllabic words. In particular, further study should investigate whether this mapping may be less regular in multisyllabic words, posing a particularly large problem for a sublexical route.

Nonword naming performance for Simulation 2 was judged by the same raters in the same manner as in Simulation 1. Nonword naming accuracy for Simulation 2 was judged to be 37.02% and 25.06%, for the Seidenberg et al. (1994) and Sibley et al. (in

press) nonwords, respectively. Of the remaining items, 46.98% and 36.67%, of incorrect pronunciations could again be corrected with a single insertion, deletion, or substitution of a phoneme. Cronbach's alpha between raters was 0.76 and 0.90, which George and Mallery's (2003) rule of thumb suggests are acceptable and excellent scores, respectively. Simulation 2 read nonwords by blending together the pronunciations for words in its lexicon. This process was akin to Glushko's (1979) propositions that reading occurs by analogy to whole word information stored in a lexicon. While Simulation 2's nonword performance is below that of a proficient reader, it demonstrates the nontrivial fact that nonword reading is possible in a model with only a single lexical route. This contradicts the primary reason to hypothesize a sublexical route, which is the supposed inability of a lexical route to read nonwords. Simulation 2 suggests a refined version may even achieve human levels of nonword reading.

As demonstrated by the DRC model, a lexical route of processing between orthography and phonology can provide a means of simulating lexical decision behaviors. Lexical routes make discrimination possible when activation passes through a locally coded lexicon. Simulation 2's lexicon nodes were implemented as radial basis functions, so each node behaved like a receptive field centered on a particular word. Words and nonwords could then be distinguished by two factors. First, exposure to a word produced full activation of the word's corresponding node. By definition, nonwords lacked a matching node in the lexicon and so only produced partially activated nodes. As a result, the level of activation on the most excited node could be used to differentiate words from nonwords. Second, English words are not distributed evenly in the space of possible

letter strings. Instead, graphemes and graphotactics tend to clump words within the possible space and so words tend to have closer neighbors than nonwords. As a result, words and nonwords produced different distributions of activation across the lexicon. In particular, words tended to have a number of orthographic neighbors that become more active than baseline. In contrast, nonwords produce flatter distributions of node activations.

Balota and Chumbley (1984) noted that lexical decision can be conceptualized as a signal detection process, where a participant differentiates word and nonword distributions along a dimension of familiarity. For Simulation 2, we created the measure,

$$fam(w) = a_1 + \sqrt{1/49 \sum_{k=2}^{50} (a_k - \bar{a})^2}, \quad (10)$$

where the first term is the activation of the most excited node in the lexicon and the second term is the standard deviation of the activations for the second through fiftieth most excited nodes. The second term stopped at the fiftieth node because including additional information did not prove useful to classifying a stimulus's lexical status. Figure 8 depicts word and nonword distributions, plotted along the axis of familiarity. These distributions and all later simulations of lexical decisions were generated with exactly the same parameter set used in simulations of word naming. Within this conceptualization, lexical decision accuracy can be simulated using a decision boundary placed along the continuum of familiarity. The average lexical decision accuracy for words and nonwords in the ELP were 87.94% and 87.97%. For Simulation 2, a boundary

parameter was fit so that 91.44% of words and 90.51% of nonwords were correctly classified.

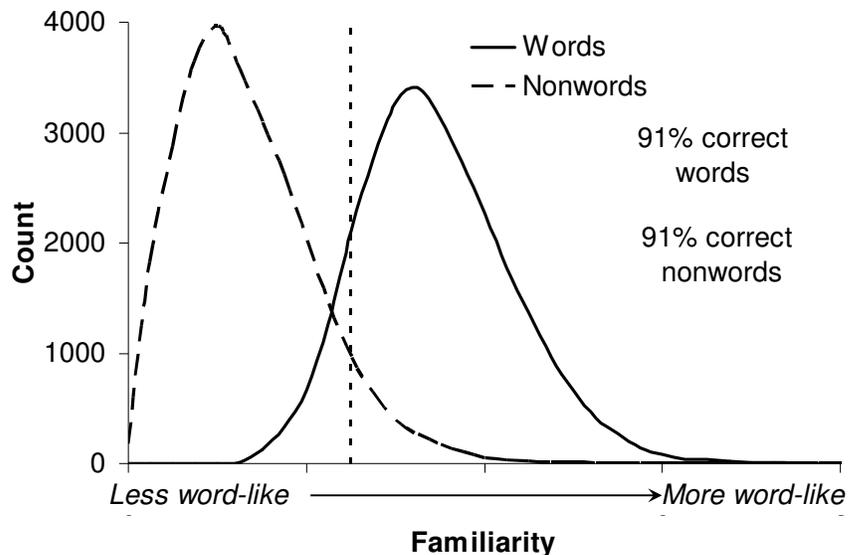


Figure 8. Word and Nonword distributions

Simulation 2 demonstrates how a model with a single lexical route can exhibit substantial capacities for word reading and recognition. Simulation 2 named far more words than Simulation 1 or any extant model. Words with corresponding behavioral data in the ELP were pronounced with a proficiency that approached that of ELP participants. Second, Simulation 2 demonstrated nonword reading abilities that approached those of Simulation 1. This comparison offers a baseline suggesting that inadequacies may result from limitations of this implementation. For instance, words included in this training

corpus that were not included in the ELP may be poor exemplar sources for analogical transfer. Or, there may be a flaw in developing orthographic wordform representations independently from learning about reading. Further, while Simulation 2's lexicon contained a large portion of an adult's vocabulary, it is certainly not full scale. As a result, Simulation 2 had fewer sources for analogical transfer than might be available to a proficient reader. Ultimately, the relatively similar nonword naming performance, between Simulations 1 and 2 suggests that a model with only a lexical route is not inherently incapable of nonword reading. This is important, as nonword reading is the behavior that generally motivates hypothesizing a sublexical route. Nonetheless, the percentage of correct nonword pronunciations again fell short of a proficient adult reader's capacities. Simulation 2 did distinguish words from nonwords at a similar rate to participants in the ELP. Later, lexical decision reaction times are calculated as a function of the distance between specific stimuli and a decision boundary between word and nonword distributions.

Taken together, these results support the viability of models with a single lexical route. It would be informative, however, to first examine a model that incorporates both sublexical and lexical routes. This dual route simulation will allow a direct comparison of the three Simulations' abilities to account for behavioral data.

## CHAPTER 4: SIMULATION 3

Simulation 3 examines the dual route approach of the currently preeminent computational models of word reading and recognition. Here, the sublexical and lexical routes from Simulations 1 and 2 were incorporated into a single model. Simulation 3 differs from current dual route architectures, in terms of its scale and how its routes are implemented. Simulation 3's sublexical route is theoretically similar to the Triangle model's direct orthography-to-phonology pathway. Simulation 3's lexical route results from passing information through a lexicon rather than semantic constraints. By virtue of localist coding, this route bears a likeness to the lexical route in DRC and CDP+.

Simulation 3 is not designed to examine how extant models may perform when scaled up to address a larger volume of data. Instead, Simulation 3 permits a direct comparison between a dual route architecture and the previously presented single route models.

### *Simulation 3: Methods*

Simulation 3 was designed to permit a controlled comparison with the other models presented herein. It used identical corpora and input/output representations to those in Simulation 1 and 2.

*Architecture.* As depicted in Figure 9, Simulation 3 was constructed analogously to Simulations 1 and 2. Sublexical and lexical routes attached Orthographic and Phonological wordform layers. Each route was constructed as in the relevant previous simulation and used the same parameters.

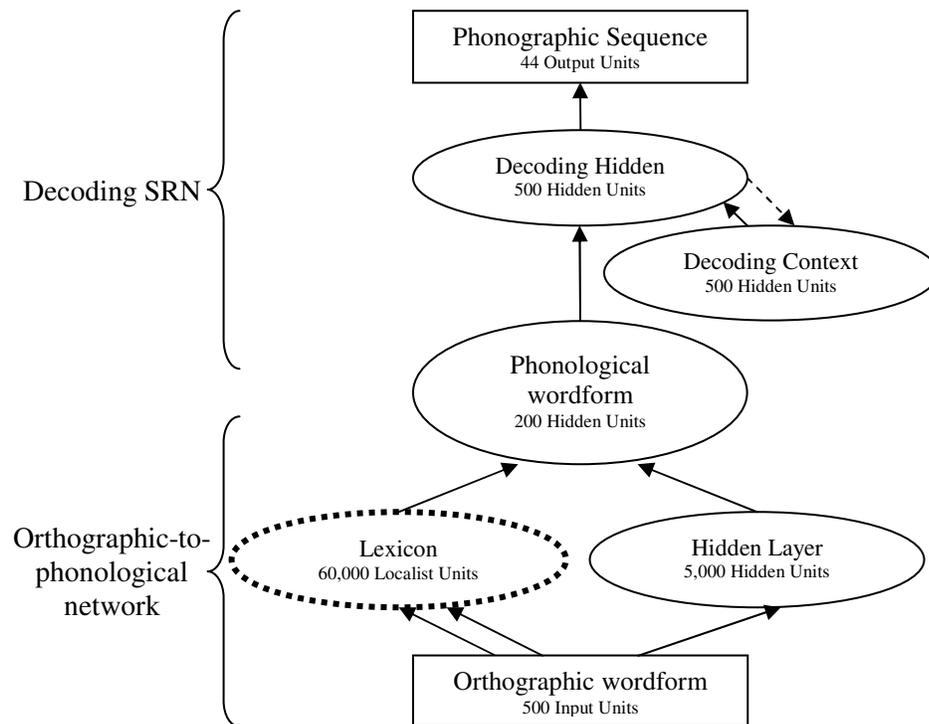


Figure 9. Architecture of Simulation 3

*Model operations.* The operations and training of Simulation 3 were identical to Simulation 1 and 2, with model training judged to asymptote after 5,500 epochs.

*Simulation 3: Results and discussion*

Simulation 3's naming performance was assessed identically to Simulation 1 & 2's and on the same corpora. Performance on the Trained ELP and Trained non-ELP words was 94.29.7% and 67.28%, respectively. Amongst the 12,059 words that Simulation 3 pronounced with at least 1 error, 79.25% of the individual phoneme and stress units were correctly produced. Of 5,120 monosyllabic words from Perry et al. (2007), 99.9% were correctly named. Again, Figure 10 illustrates that phonemic length strongly influenced which items were named correctly.

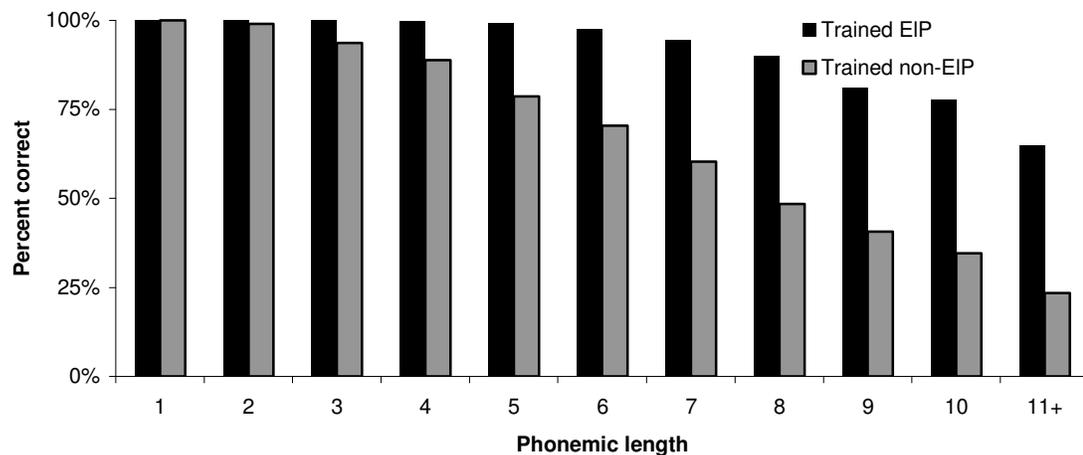


Figure 10. Naming accuracy for Simulation 3, as a function of length

Simulation 3 had very similar word naming performance to Simulation 2. To ensure Simulation 3's naming performance was not entirely attributable to its lexical route, its sublexical route was removed and phonemic outputs regenerated. Without its sublexical route, Simulation 3 only correctly named 590 words in its 60,000 items training corpus, 81.69% of which were monosyllabic. Of the items with at least 1 error, 39.99% of the correct phoneme and stress units were generated. This performance decrement suggests that both lexical and sublexical routes contributed to generating phonemic outputs.

Nonword naming performance for Simulation 3 was judged as in Simulation 1 and 2 and by the same raters. Simulation 3's nonword naming accuracy was judged to be 50.85% and 32.11%, for the Seidenberg et al. (1994) and Sibley et al. (in press) nonwords, respectively. Again, 38.10% and 34.11% of incorrect pronunciations could be corrected with a single phoneme's insertion, deletion, or substitution. Cronbach's alpha between raters was 0.914 and 0.894, which George and Mallery's (2003) rule of thumb suggests are excellent and good scores. Again, this nonword reading performance falls short a proficient reader's capacities.

Simulation 3 read nonwords with slightly better accuracy than Simulation 2, but worse accuracy than Simulation 1. At the least, this demonstrates that dual route models do not necessarily inherit the best abilities from each of their routes. Simulation 3's performance may reflect shortcomings in this particular instantiation. It is probable that nonword naming performance would have increased through tuning parameters on the two routes. In the limit, manipulating parameters could presumably make this dual route

model read words at least as well as either single route model. In contrast, our approach was to use the same relevant parameter sets across all simulations.

Simulation 1 demonstrated our sublexical route was incapable of supporting lexical decision performance. So we adopted the approach of DRC and CDP+ where word recognition was achieved exclusively by the lexical route. As connections between orthography and the lexicon were assigned by identical algorithms in Simulations 2 and 3, the two models had identical lexical decision performance.

The dichotomy between sublexical and lexical knowledge emerges behaviorally in the tasks of word reading and recognition. The Triangle, DRC, and CDP+ models architecturally instantiate this dichotomy in two routes between orthography and phonology. This offers a means of avoiding contrasting requirements between reading generalization and distinguishing known from novel stimuli. However, this approach is predicated on a best of both worlds scenario, where the model inherits the best attributes of each route. In contrast, Simulation 3 blended the capacities of sublexical and lexical routes, demonstrating that dual route models do not necessarily offer capacities beyond single route models. But before drawing final conclusions, we examine all three of our simulations abilities to account for behavioral data.

## CHAPTER 5: BEHAVIORAL DATA

The Triangle, DRC, and CDP+ models were designed to account for the results of factorial experiments (i.e., Perry et al., 2007). Generally speaking, these types of studies demonstrate that reading and recognition latencies differ on small groups of monosyllabic words which embody a stimuli characteristic's extremes. However, Balota et al. (2004) showed that computational models can emulate numerous factorial studies while accounting for fairly little variance at the item level. Addressing item-variance is an essential test of a model as it reflects the ability to simulate a behavior rather than isolated behavioral effects. So we first report each simulation's capacity to address item-variance in naming and lexical decision latencies for roughly 28,000 English words. Then we report specific variables effects upon behavioral and simulated latencies. We do not address the results of particular factorial studies, which focus on small numbers of monosyllabic words. Instead, our Simulations are used to predict each variable's effect size and direction on human latencies regarding more than 7,500 mono and multisyllabic words.

### *Item level analyses*

Word naming latencies were simulated as the uncertainty in model outputs. Uncertainty is used as a proxy for reaction times in connectionist networks without dynamics, where articulatory processes are hypothetically delayed by indefinite outputs (Plaut et al., 1996). Specifically, the naming reaction time for each word,  $w$ , was calculated as,

$$NameRT(w) = \left( \sum_{t=1}^{Phonemes} 1 - a_{Most}^{[t]} \right)^X, \quad (12)$$

where  $a_{Most}^{[t]}$  is the activation of the most energized output unit in each bank, on each time step,  $t$ . The parameter  $X$  was set to produce a linear fit between each model's predictions and behavioral data in the ELP. The free parameter  $X$  was set to 0.4 for Simulation 1 and 0.2 for Simulations 2 and 3.

The bottom row of Table 1 displays the amount of item-variance that each of our simulations address in ELP naming latencies for 28,032 words. As extant models can not address this dataset, additional rows in table 1 report item analyses for intersections of our corpus with their training corpora. Intersecting our corpus with the corpus from CDP+ left 5,120 monosyllabic words. A third stimuli set with 2,730 monosyllabic words was generated by additional intersections of the DRC and Triangle models' corpora. The behavioral data utilized is all comparisons were provided by the ELP. For item level analyses of naming accuracy see Appendix 2.

Table 1. Item-variance accounted for by various models on different stimuli sets

Intersected corpora	# items	Percent of item-variance accounted for by:					
		Triangle	DRC	CDP+	Sim. 1	Sim. 2	Sim. 3
Triangle, DRC, CDP+, Sim. 1, 2, 3	2,730	2.9	6.0	18.0	17.3	12.8	15.9
CDP+, Sim. 1, 2, 3	5,120	-	-	18.8	18.6	15.6	18.7
Sim. 1, 2, 3	28,032	-	-	-	34.2	33.9	35.4

Simulations 1, 2, and 3 addressed an unprecedented quantity of item-variance in 28,032 mono- and multisyllabic English words. These three simulations accounted for similar amounts of item-variance. This suggests they are sensitive to similar underlying variables, regardless of the processing mechanisms they utilize. Later analyses reinforce this claim. This shared sensitivity presumably arises from the models utilizing similar representational and learning systems.

CDP+ and the current simulations account for similar quantities of item-variance in naming latencies for monosyllabic words. So this work demonstrates that computational models can be scaled up to address multisyllabic word reading, without compromising their ability to address monosyllabic naming. The percentage of variance that our simulations address in the monosyllabic corpora is roughly half of the item-variance addressed in the larger dataset. This was because monosyllabic words represent a restricted range of characteristics like length and frequency, which strongly influence behavioral and simulated latencies. Effectively, a larger portion of variance in monosyllabic reading latencies results from individual differences (Yap, 2007), which our models do not address.

Lexical decision latencies were not created for Simulation 1, as it could not distinguish words from nonwords. The lexicon used in Simulations 2 and 3 provided a proxy of lexical decision latencies,

$$LexDecRT(w) = -|fam(w) - crit|, \quad (13)$$

where  $fam(w)$  is the stimuli's familiarity calculated according to equation 10,  $crit$  is a decision boundary between word and nonword distributions.  $Crit$  was set to 2.5 so that the 91.44% of the words and 90.51% of the nonwords were correctly classified. This is similar to data in the ELP, where average naming accuracy for words and nonwords was 87.94% and 87.97%, respectively. Effectively, recognition latencies were calculated as the distance between a stimuli's familiarity and a decision boundary, as suggested by Balota and Chumbley (1984). Simulations 2 and 3 accounted for 35.64% of the variance in lexical decision reaction times for 27,881 words. This metric was generated exclusively from lexicon activation, which does not incorporate information about the serial process that activates the lexicon. Thus, it does not approximate how stimuli with more letters tend to take longer to visually input. This effect can be approximated by adding a second term to equation 13 to create,

$$LexDecRT(w) = -|fam(w) - crit| + l(w) / 4, \quad (14)$$

where  $l(w)$  is the orthographic length of word  $w$ . Incorporating a length term allows Simulation 2 and 3 to account for 41.56% of the item-variance in reaction times to words. To our knowledge, the Triangle, DRC, and CDP+ models have not been used to address item level lexical decision latencies. As a result, direct comparisons between current and extant simulations were not possible.

### *Word characteristics' impact on reaction times*

Next, we examine each simulation's sensitivity to stimuli characteristics that affect reading and recognition behaviors. Perry et al. (2007) used CDP+ to replicate a set of benchmark factorial studies that analyzed reading latencies for small numbers of monosyllabic words. In contrast, we examine how these variables affect latencies for thousands of mono and multisyllabic words that reflect a large range of each variable. Further, we analyzed all of these variables effects upon reading as well as recognition latencies.

Yap (2007) examined how numerous stimuli characteristics affected the naming and lexical decision latencies for 9,639 words in the ELP. Roughly 7,500 (7,531 for naming, 7,432 for lexical decision) of these words also appeared in the current simulations' training corpus. The words in this stimulus set had between 1 and 5 syllables, with an average of 1.75 syllables. With these stimuli we examined the relative effect sizes and directions between theoretically important variables and human or simulated latencies. Comparing effect sizes provides a graded means of assessing each simulation's ability to replicate behavioral effects. This is in contrast to replicating factorial studies, where primary importance is given to statistically significant differences in behaviors regarding small stimuli groups.

There is substantial multicollinearity in the variables that impact behavioral latencies. In some cases, variables measure different aspects of a class of word characteristics. For instance, the number of phonemes and syllables in a word are

obviously related. In other cases, multicollinearity results from word characteristics being confounded in English. Some of these two way interactions have become theoretically important and so we also report these effects. Interactions were assessed as suggested by Cohen et al (2003) and reiterated by Yap (2007), the two relevant main effect variables were first entered into a stepwise regression and then their interaction was entered. The differences in the  $R^2$  for the first and second steps are reported as effect sizes.

Table 2 displays  $R^2$  between word characteristics and naming latencies. Table 3 displays analogous information for lexical decision latencies. All of the effects displayed in these tables are in the correct directions and statistically significant, unless denoted by an asterisk. In particular, the three interaction variables were in different directions for lexical decision latencies.

Table 2. Analysis of  $R^2$  in naming latencies

	ELP	Sim. 1	Sim. 2	Sim. 3
Frequency	.309	.212	.154	.151
Phonological length	.279	.497	.255	.284
Syllabic length	.249	.436	.289	.270
Syllabic after Phonological Length	.020	.034	.052	.021
Coltheart's N	.178	.162	.056	.063
Levenshtein orthographic Distance	.366	.628	.336	.372
Levenshtein consistency	.146	.205	.088	.090
Stress typicality	.153	.258	.117	.117
Frequency * Orthographic length	.044	.104	.121	.184
Frequency * Levenshtein Orthographic Distance	.032	.074	.114	.196
Frequency * Levenshtein consistency	.024	.046	.063	.067
Correlation with ELP effect sizes		.87	.78	.67

The correlation between effect sizes for ELP participants and each simulation reflects each simulation's ability to predict the effect of stimuli characteristic upon reading and recognition. With respect to naming latencies, Simulation 1 had the highest correlation and Simulation 3 the lowest. Contrary to a number of claims that will be discussed, this shows that none of these particular phenomena necessitate hypothesizing dual routes in the reading system. With respect to lexical decisions, Simulations 2 and 3 strongly predict the magnitude of each effect. This correlation is effectively unchanged (0.984 vs. 0.976) if the three interaction variables are excluded.

Table 3. Analysis of  $R^2$  in lexical decision latencies

	ELP	Sim. 2 & 3
Frequency	.427	.806
Phonological length	.183	.415
Syllabic length	.205	.346
Syllabic after Phonological Length	.032	.011
Coltheart's N	.113	.296
Levenshtein orthographic Distance	.257	.456
Levenshtein consistency	.089	.113
Stress typicality	.114	.195
Frequency * Orthographic length	.021*	.001*
Frequency * Levenshtein Orthographic Distance	.010*	.006*
Frequency * Levenshtein consistency	.009*	.000*
Correlation with ELP effect sizes		.98

One of the most robust findings in this field is that words which are encountered more often tend to be read and recognized faster (Forster & Chambers, 1973). The logarithm of a word's frequency is one of the strongest predictors of its naming and lexical decision latencies, whether the word is monosyllabic (Balota et al., 2004) or multisyllabic (Yap, 2007). As a result, frequency has played a substantial role in theorizing about the lexical system. For instance, Forster & Chambers (1973) argued that lexical access involves a serial search through a list ordered from the most to least frequent word. In contrast, the Triangle model and our Simulations exhibit this effect due

to frequency based scaling of the error terms that drives learning. This error scaling approximates how words that are encountered more often have more opportunities to be learned. DRC, CDP+, and Simulations 2 and 3 exhibit a frequency effect in lexical decision latencies, because of a multiplicative frequency term placed on the activation functions of lexicon nodes. This is akin to a proposition by Morton (1969) where frequency multiplicatively scaled a threshold on word detectors. In contrast, lexicon nodes in our models produced graded responses rather than binary distinctions according to decision boundaries.

Another theoretically significant finding is that shorter words tend to be read and recognized faster than longer words (Weekes, 1997; Ziegler, Perry, Jacobs, & Braun, 2001). Coltheart et al. (2001) argued that length effects indicates a serial processes involvement in reading, which is supposedly contrary to a fully parallel architecture like the Triangle model. Yet Simulation 1 adopts only standard parallel distributed processing mechanisms, but incorporates a serial aspect via the simple recurrent networks involved in orthographic and phonological representation. As a result, Simulation 1 displays a strong effect of phonemic length upon naming latencies. Simulations 2 and 3 exhibit phonological length effects upon naming and lexical decision latencies.

ELP participants and all of our simulations exhibit an effect of syllabic length. These effects persist when variance attributable to phonological length is partialled out. To our knowledge models of word reading and recognition have never addressed this effect. New, Ferrand, Pallier, & Brysbaert (2006) argued that the behavioral impact of syllabic length implies a functional role of the syllable in the cognitive system, but our

simulations do not explicitly represent or employ syllables. It is possible that they exploited syllabic dependencies by implicitly coding syllabic information into connection weight matrices. There may also be a more transparent explanation; this effect upon simulated naming latencies may just reflect the number of vowels in a word. Vowels tend to embody more complex orthographic-to-phonological mappings than consonants and so our simulations tend to activate vowel output units less strongly than consonant units. This could explain the observed sensitivity to syllabic length, as syllables are delimited by vowels. This begs the question, of whether simulating this effect is serendipitous or if participant's sensitivity to syllabic length is also a function of the number of vowels.

Numerous behavioral experiments have explored how a given word's processing is affected by similarly spelt words (Coltheart et al., 1978; Glushko, 1979; Ziegler et al., 2001). One way to operationalize this effect is with orthographic neighborhood size, which reflects the density of words within regions of orthographic space. An emerging consensus is that large numbers of orthographic neighbors facilitate naming and lexical decision (for a review, see Andrews, 1997). The most utilized measure of neighborhood size is Coltheart's N, which for a given word denotes how many other English words can be created by substituting a single letter. Coltheart's N operates well for monosyllabic words where orthographic neighbors are generally of the same length. With respect to longer words, seemingly similar spelt stimuli are often of different lengths. So a multisyllabic word's neighbors should be defined by letter substitutions, as well as additions and deletions. Yarkoni et al. (submitted) proposed such a metric called

Levenshtein orthographic distance, which is computed as the average of the minimum number of letters that must be added, removed, or substituted to transform a word into its nearest 20 orthographic neighbors. ELP participants and Simulations 1, 2, and 3 all demonstrate statistically significant effects of Coltheart's N and Levenshtein orthographic distance. Effect sizes for human and simulated latencies are larger for Levenshtein orthographic distance, which suggests it is the superior metric.

The degree to which a word's pronunciation follows common grapheme-to-phoneme correspondences is referred to as its regularity (Seidenberg, Waters, Barnes, & Tanenhaus, 1984). In contrast, a word's consistency reflects whether it is pronounced similarly to orthographically similar words (Glushko, 1979). There is substantial debate about the roles of regularity versus consistency; however, using carefully controlled stimuli, Jared (2002) reported that naming latencies were influenced by the consistency of a word's pronunciation regardless of its regularity. To define consistency with respect to mono- and multisyllabic words we utilized Levenshtein consistency (Yarkoni, Balota, & Yap, submitted). Levenshtein consistency was calculated as the ratio of a word's Levenshtein orthographic distance to its Levenshtein phonological distance. These terms are computed as the average of the minimum number of letters or phonemes that must be added, removed, or substituted to transform a word into its nearest 20 neighbors. This ratio diverges from 1, when a word's orthographic and phonological neighbors represent different underlying words. This indicates an inconsistent mapping. ELP participants and all of the current Simulations demonstrated that naming latencies were facilitated by consistent orthographic-to-phonological mappings. Simulation 1, 2, and 3's sensitivity to

consistency emerged because training drove learned connection weights to extract statistical relationships across words in the corpus.

The assignment of vowel stress can only be meaningfully simulated in models that read multisyllabic words. According to Rastle & Coltheart (2000) 83% of disyllabic English words receive first syllable stress. Sereno (1986) examined disyllabic words' stress patterns as a function of part of speech and found that 93% of nouns receive first syllable stress, in contrast to 76% of verbs. This prompted Rastle & Coltheart's (2000) creation of a complex algorithm that computes stress assignment. This algorithm operates at a lexical level, as sublexical information is not easily related to part of speech. Yet Simulations 1, 2, and 3 all exhibit effects of stress typicality which mimic the effect observed in ELP participants. This shows it is not necessary to hypothesize additional stress assignment mechanisms specifically to address multisyllabic word reading and recognition. Further, Simulation 1's sensitivity to the stress typicality effect calls into question whether stress assignment must operate on a lexical level.

Three interactions of the previously discussed variables are considered as theoretical important. The interactions of frequency with length, orthographic neighborhood, and consistency have been interpreted as evidence for dual route theories (Coltheart et al., 2001). These arguments begin with the assertion that higher frequency words are supported by a lexical route, while lower frequency words are supported by a sublexical route. DRC construes a lexical process as operating in parallel, while a sublexical process operates serially. This would produce a frequency by length interaction, where serial sublexical processes are more pronounced for low frequency

words. Further, DRC's lexical route embodies inhibitory interactions between lexical nodes, while its sublexical route operates most efficiently upon regions of orthographic space which are densely populated. This has been used to explain why the interactions of frequency and Levenshtein orthographic distance impacts naming latencies. Finally, Coltheart et al. (2001) suggest that the reading of higher frequency words is supported by a lexical route, which is not sensitive to consistency in the mappings between letters and phonemes. In contrast, a sublexical route would be sensitive to consistency. This has been used to explain the observed interaction of frequency and Levenshtein frequency. Simulations 1, 2, and 3 all display effects of these three interactions upon naming latencies. This challenges whether these interactions are evidence for dual route theories.

## CHAPTER 6: GENERAL DISCUSSION

This dissertation's goal was to scale up computational models of word reading and recognition, while exploring the necessity of theorizing two distinct computational routes. We presented three simulations that each processed an order of magnitude more words than any previous model. This was achieved via representational systems that enabled the processing of mono- and multisyllabic words. By systematically varying the models' architectures, we explored the necessity of hypothesizing two distinct routes in the reading system. Simulations 1 and 2 utilized either a single sublexical or lexical route, while Simulation 3 incorporated both routes.

We first explored each simulation's capacity for word reading and recognition. We focused on these abilities, because reading of nonwords is said to rely on sublexical processes, while word recognition supposedly requires a lexical process (i.e., Coltheart et al., 2001; Perry et al., 2007). The Triangle, DRC, and CDP+ models propose distinct sublexical and lexical routes, to cope with the opposing requirements of these tasks. In particular, nonword reading seemingly requires treating words and nonwords similarly, while lexical decision tasks require distinguishing known from novel stimuli.

Simulations 1 and 2 performed word reading and recognition tasks with a single sublexical or lexical route, respectively. Simulation 1 confirmed that a sublexical route is

poorly suited to word recognition. Simulation 2 suggested that a single lexical route may be sufficient for both reading and recognition. Simulation 2's nonword reading performance was below a proficient reader's abilities, but nevertheless a large number of mono- and multisyllabic nonwords were acceptably pronounced. This challenges the assertion that a lexical route is fundamentally incapable of nonword naming and instead suggests a refined version may achieve an adult's levels of performance. Simulation 3 combined both sublexical and lexical routes from the previous simulations. Analyses of Simulation 3 demonstrated that dual routes provide negligible benefits over a single lexical route. This begs the question of whether additional assumptions inherent in a dual route system are warranted.

Simulations 1, 2, and 3 addressed substantial amounts of behavioral data in the ELP. Each simulation accounted for roughly double the item-variance of any previous model, in naming latencies for mono- and multisyllabic words. For a direct comparison with previous work, we intersected our corpus with the corpora used by the Triangle, DRC, and CDP+ models. Regarding only monosyllabic words, our simulations accounted for more variance than the Triangle or DRC models and performed similarly to CDP+. In item analyses of lexical decision latencies, Simulations 2 and 3 accounted for an unprecedented amount of variance, while Simulation 1 was incapable of mimicking this behavior. To our knowledge, the Triangle, DRC, and CDP+ models have not been used to address item-variance in this behavior.

As discussed by Seidenberg and Plaut (2006) the goal of computational modeling is not to fit behavioral data, but rather to inform theorizing about the cognitive system.

So models of reading should address item-variance and elucidate how single word reading and recognition are affected by theoretically important variables. To this end, we examined each simulation's sensitivity to stimuli characteristics that impact behavioral latencies. Our simulations mimicked all of these variable's effects upon ELP participants' latencies, for mono- and multisyllabic words. There was a strong correspondence in each variable's effect size for human and simulated latencies. Notably, each model emulated behavioral phenomena specific to multisyllabic word reading, such as syllabic length and stress typicality. This suggests that computational models can be scaled up to address multisyllabic reading, without incorporating numerous new mechanisms specifically for this purpose. This is contrary to previous work which proposed that syllables needed a privileged position in the reading system (Cutler, Mehler, Norris, & Sequi, 1986) or that an additional algorithm was necessary for stress assignment (Rastle & Coltheart, 2000). Finally, analyses of naming latencies showed that three stimuli interactions, which are considered as evidence for dual route architectures were simulated in both of our single route models.

One motivation for single route models stems from the observation that learning to read scaffolds upon prior learning about speech (Kello, 2006). This implies a developmental model should first acquire associations between phonology and semantics and then link orthographic inputs with the intermediate representations between phonology and semantics. This rationale motivated the Junction Theory, which is schematically depicted in Figure 11 (Kello, 2006). Here, orthographic, phonological, and semantic layers of representation are connected via a junction. This single route provides

the only means of interaction between orthographic, phonological, and semantics representations.

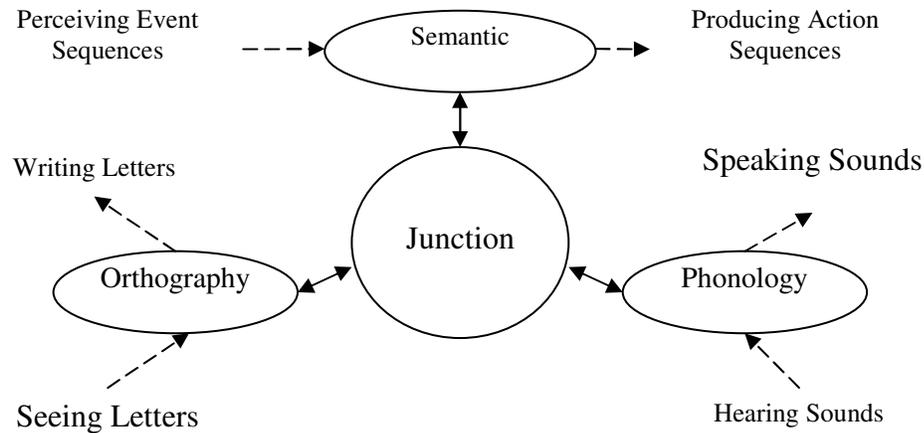


Figure 11: The Junction Theory of word reading and recognition

Simulation 1 challenges the feasibility of a sublexical Junction Layer. In all natural languages the mapping between phonology and semantics is largely arbitrary (Hockett, 1960). Morphemes produce some systematicity in sublexical relationships between phonology and semantics, but this regularity is relatively weak. Learning this mapping with connectionist machinery would produce distributed hidden representations that embody fairly little sublexical information. Learning to read would then require solving a very complex mapping between orthography with these idiosyncratic junction

representations. Further, if semantic influences on the junction layer are strong enough to simulate lexical decision behaviors, then distributed junction representations for known and novel stimuli would in all likelihood be too dissimilar to support nonword naming.

In contrast, Simulation 2 indicates the plausibility of a Junction Layer that takes the form of a lexicon. This lexicon would constitute a single lexical route between orthography, phonology, and semantics. While others have argued that a lexical route could not support nonword naming (i.e., Coltheart et al., 2001) Simulation 2 challenges this claim. Architecturally distinguishing words from nonwords with a lexicon provides a means to achieving word recognition. Naming behaviors could still occur through a process of analogy to known word's pronunciations, as proposed by Glushko (1979).

By definition a lexicon utilizes localist representations, but it does not necessarily utilize localist processing. As a parallel case [pun intended] Page (2000) noted that Rumelhart and McClelland's (1981) Interactive Activation model utilizes localist representations and yet it is a seminal work in the PDP literature. Each word processed by an Interactive Activation network is coded locally, but multiple nodes become active in response to a single input. As a result, multiple word nodes participate in the parallel distributed processing of a single input. Similarly, orientation columns in visual cortex represent information locally; each column represents a line, with an orientation, in a particular receptive field. However, visually processing the text of a single word involves parallel and distributed processing across many orientation columns. The lexicon used within Simulation 2's lexical route operates analogously. Each node instantiates a receptive field centered on a particular word, yet nodes produce graded

responses to a class of stimuli. As a result, reading is supported by sparsely distributed patterns of activation across the lexicon.

### *Limitations and extensions*

The simulations presented in this dissertation did not address every phenomenon related to single word reading and recognition. Instead, our framework was designed to be extended to address additional findings. First, as with the Triangle, DRC, and CDP+ models, we did not fully implement semantics. Yet relating semantics to orthography and phonology is the primary goal of the language system (Hockett, 1960). Concretely, semantic impacts on reading are made apparent in homographs, where meaning must disambiguate between the noun and verb pronunciations of DOVE. Semantics could be incorporated into our current models, using representations developed by Rohde (2007) or Jones and Mewhort (2007). Harm and Seidenberg (2004) and later Kello (2006) demonstrated the feasibility of incorporating these kinds of representations into models using connectionist machinery. By implementing semantics our models would be capable of simulating semantic variables' impact on word reading and recognition, in particular imageability (Cortese & Fugett, 2004), number of semantic associates (Toglia & Battig, 1978), number of semantic senses (Miller, 1990), and semantic priming (Plaut & Booth, 2000).

A complete model of the reading system should address the dynamics of interactions amongst its components. Behavioral evidence suggests the presence of reciprocal effects in the reading system. In the current simulations activation always

flows from orthography towards phonology. This is not a theoretical commitment, but rather a useful simplification while we focus on other issues. Nonetheless, behavioral naming latencies are impacted by the consistency in the mapping from phonology back to orthography (Stone, Vanhoy, & Van Orden, 1997). This feedback process becomes apparent in the case of homophones, i.e., YOKE and YOLK, whose naming latencies systematically differ from control words. Incorporating reciprocal or bidirectional connectivity into our models would permit simulations of effects like feedback consistency and would increase the models fidelity to the Junction theory. Eventually, this may create a single route model capable of simulating any behavior that involves mapping between the orthographic, phonological, and semantic forms of a word. Such a model could read and recognize words, but also perform transcription, phonological shadowing, and numerous other behaviors.

The present simulations were designed to simulate the behaviors of proficient adult readers. As a result, they do not directly address a large body of phenomena regarding reading acquisition (Gleason, 1951; Resnick, Lauren, Phyllis, & Weaver, 1976; Cunningham & Stanovich, 1991; Saffran, Aslin, & Newport, 1996; Dufva, Niemi, & Voeten, 2001; Ziegler & Goswami, 2005; Goswami & Ziegler, 2006). However, our models incorporate connectionist learning algorithms and so could be extended to address this domain. The only components of our simulations that were not learned were the lexicons employed by Simulations 2 and 3. A lexicon, which utilizes localist representations, can not be created by traditional PDP learning algorithms, like back propagation of error. However, other learning algorithms can develop localist

representations and these algorithms could be adapted to our purpose (Grossberg, 1980). Simulating normative reading acquisition is a necessary step before simulating developmental dyslexias, which are a major source of evidence in hypothesizing about the reading system (Marshall & Newcombe, 1966; Graham, Hodges, & Patterson, 1994; Schwartz, Marin, & Saffran, 1979; Castles & Coltheart, 1993; Coltheart, Patterson, & Marshall, 1980; Shallice & Warrington, 1980; Patterson, 1982; Plaut, Behrmann, Patterson, & McClelland, 1993; Plaut et al., 1996).

The current models are intended to set the groundwork for simulations of two prevalent forms of acquired dyslexia. Acquired surface and phonological dyslexia have been interpreted as selective damage to sublexical or lexical routes (Coltheart et al., 1993). However, these deficits could instead reflect excessive application of sublexical or lexical information. This theory would not imply the presence of separable routes that embody a hard sublexical-lexical distinction. In the current single route models, particularly Simulation 2, acquired dyslexias could be produced by varying a control parameter that modulates the size of lexicon node's receptive fields. By increasing the size of receptive fields, aberrantly large numbers of words would influence the processing of a given stimuli. In effect, large numbers of exemplars would drive processing to conform to majority orthographic to phonological mappings, rendering words with irregular mappings unpronounceable. Decreasing the size of receptive fields would cause each stimulus to be processed with respect to a single known exemplar. This would render nonword reading impossible. Within a single-route framework acquired dyslexias could be interpreted as stable modes of interaction that give rise to

characteristic modes of processing. An unimpaired lexical system could flexibly shift among modes in response to task demands. However, damage to cortex could limit the modes available to the system. In the extreme, the reading system would become locked into a particular mode where a selective impairment is observed. Within this account, acquired surface and phonological dyslexia could represent reading systems locked into lexical and sublexical modes, respectively.

These and related issues will be explored in future work.

## APPENDIX 1: THE EFFECT OF DEPENDENCIES

We discuss the structure of word forms in terms of the dependencies that occur among the sequence's elements and positions. It has been shown that both infants and adults can implicitly learn about the dependencies in linguistic sequences (Aslin, Saffran, & Newport, 1999; Jusczyk, Luce, & Charles-Luce, 1994; Saffran, 2001; Saffran, Aslin, & Newport, 1996). Strings of elements that conform to these learned dependencies can be considered as legal with respect to that language, while strings embodying lower levels of conformity are less legal. It is important to note that the dependencies among letters and phonemes are too probabilistic to strictly separate legal from illegal word forms. Instead they provide soft constraints that can be formalized as a probabilistic grammar (Jurafsky, 2003; Seidenberg & MacDonald, 1999).

This graded legality distinction has been shown to impact language user's performance on a number of lexical tasks. For instance, participants can identify single letters more rapidly when they are embedded within words relative to random sequences of letters (Reicher, 1969; Bergman, Hudson, & Eling, 1988; Frisch, Large, Zawaydeh, & Pisoni, 2001; Grainger, Bouttevin, Truc, Bastien, & Ziegler, 2003; Maris, 2002; Wheeler, 1970). This effect respects the graded quality of legality; with single letters being identified faster in more legal nonwords than in less legal ones (Aderman & Smith, 1975; Baron & Thurston, 1973). Legality also impacts numerous phonological tasks; words tend to be read faster when they end in high, relative to low, frequency syllables (Levelt

& Wheeldon, 1994) and legal strings of phonemes tend to be recognized faster and more accurately than less legal ones (Brown & Hildum, 1956; Vitevitch & Luce, 1999).

The most direct evidence for language user's sensitivity to legality is an individual's ability to distinguish between sequences that do and do not conform to the statistical regularities of their language. Bailey and Hahn (2001) showed that subject's judgments of a nonword's legality could be predicted in part by the nonwords conformity to transitional probabilities in English words. This predicative power was largely independent of the nonwords similarity to particular English words, with measures of the nonwords lexical neighborhood accounting for a largely independent portion of variance in the subjects' judgments.

This sensitivity to legality appears to originate at very early stages of development; Jusczyk et al., (1993) demonstrated that infants attend longer to nonwords with common phonotactic patterns than nonwords with less common patterns. Some researchers suggest that cortical structures are responsible for learning about phonological legality (Auer & Luce, 2003). Functional neuroscientific techniques have corroborated this idea, with studies showing differential electrophysiological responses originating in superior temporal cortex as a function of legality (Bonte et al., 2005).

## APPENDIX 2: ITEM-VARIANCE IN NAMING ACCURACY

To better examine our Simulation naming performance we assessed whether each model could predict item-variance in word naming accuracies from the ELP database. To our knowledge, item-variance in naming accuracy has not been used to assess prior lexical models, where the focus is on predicting reaction times. Nonetheless, the ELP provides a large volume of data regarding naming accuracy and so this behavior should be used to assess new models. Addressing this data requires a more sensitive measure than the binary distinction employed thus far. A graded measure of naming accuracy was created by determining how far each output was from the target sequence. Specifically, word naming accuracies were calculated as,

$$NameACC(w) = \prod_{t=1}^{Phonemes} a_{Target}^{[t]}, \quad (15)$$

where  $a_{Target}^{[t]}$  was the activation of the target output unit in each bank on each time step. The activation on each output unit can be interpreted as the probability of producing the corresponding output, as a result of the divergence error term used to train the model. The product of these activations reflects the probability of correctly producing each output. This measure deviates from the proxy for naming reaction time, which was a function of the maximally active units in each bank of outputs. Regressing  $NameACC(w)$  onto the naming accuracies provided in the ELP, Simulation 1, 2, and 3 accounted for 6.51%, 8.12%, and 8.43% of the item-variance in naming accuracies for the 28,032 Trained ELP words. Models should be expected to predict less variance in measures of

accuracy than in reaction time, as the variables known to impact these behaviors account for more variance in measure of naming latency (Balota et al., 2004).

### APPENDIX 3: NONWORD CORPORA

From Seidenberg et al. (1994):

baint, bange, barce, barsh, bartz, baugh, beese, beil, beint, belf, belm, bense, bibe, bierce, biff, bimpse, binc, binch, bint, bip, bipe, blaft, blan, blush, blaunt, blea, blex, blypt, boach, boaf, boarse, boist, bonge, borge, bort, bove, braist, braze, breat, brewn, brild, brist, brune, buch, bup, burf, bux, byth, cack, chamb, chank, chape, chaut, chawn, chazz, chence, cherd, chig, chis, choad, choll, chone, chor, chure, chye, clart, cleash, clert, cles, cleve, clise, clo, cloor, clurt, clyle, cooze, cound, craid, crame, creal, crean, creet, creighth, crelt, crent, cryke, cuce, curnt, dacht, dade, dafe, dain, dar, dask, daste, dath, datt, dench, denth, derch, dewt, diend, dieve, dilge, dilt, dirm, disp, dithe, dixth, doath, dode, doir, dold, doof, doup, draille, drang, drase, dre, drebb, dreer, drel, drept, drit, droap, drock, dront, drook, drost, drouth, drow, druile, drust, duess, duge, duilt, dur, durb, durse, dush, dyst, fache, fane, fard, faunch, fauze, feant, feap, feath, feece, fey, fich, fieck, finth, fipt, firch, firk, fiss, fize, flas, fleik, floth, flun, flutch, foast, fod, fonce, fong, fooch, foon, foose, forch, foun, founce, frad, frand, frast, freamt, frell, fren, frewe, froke, fru, fruise, fruke, fruse, fulse, fumb, furk, garge, gat, gerse, gick, gieze, gign, glab, glarc, glay, gleath, gleard, glebt, glep, glesh, glief, glithe, glourt, gluff, glusk, goak, goise, golk, gomb, gou, grall, graw, graxe, greep, greft, grend, grood, groot, gruite, gruy, gulb, habe, halm, hapt, heam, hease, hef, hegg, hene, henth, hepth, herf, herge, hifth, hile, hilk, hine, hink, hisk, hoat, hodd, hoost, hoothe, horst, hosh, howd, huild, hulp, jump, jate, jaucee, jealm, jeri, jide, jind, jir, jitt, jom, jope, jore, jum, jyre, kag, kail, karch, keace, kearne, kess, kie, kirth, kith, konze, koust, kurst, larb, larf, leige, lext, lirage, litz, loice, lorce,

lorm, lourse, lourth, ludge, luzz, mact, mage, malve, manch, mangst, marn, masp, meave, meeve, merid, meize, melch, melfth, mier, mim, mird, mirst, mish, miz, moax, modge, moft, moid, molf, motch, mounge, mource, mourd, moy, moest, mulge, mulk, murd, myp, nadd, naise, narve, nault, neak, neld, nerr, nerth, nid, neist, nilth, ninx, nire, moil, nooth, norld, nounge, nouse, nowth, nube, nuck, nurge, padge, paff, pait, paik, pawk, ped, pelve, peme, pern, peud, pice, pidst, piege, pift, pight, plaive, plap, pleasr, plerk, plew, plewd, pling, plon, ploop, plourn, plown, ploxx, plue, plut, pog, poin, poove, poss, pral, preadth, preel, prot, pude, puit, pult, pung, purn, pute, pymm, raim, ralp, rance, rause, rawl, reast, redge, reeze, relte, rem, ret, ri, rield, riew, rike, ril, ringe, rive, rix, ronk, roo, roosh, rould, rounce, rud, rull, rund, rynch, saisle, sanse, scole, scra, seaf, seart, seb, sempt, sheapt, shearse, shelk, sial, sib, sidth, silm, simb, skose, slere, slote, smair, samise, smapse, smaught, smead, smein, smill, smough, smoul, smuard, smuice, smuide, snass, snaube, sneed, sneue, snoam, snoud, snurr, sny, sobe, somp, sond, sorn, sorth, sount, spake, spathe, spauge, spaul, speight, spetch, spoan, spowl, spram, staltz, stamn, starp, steach, stearth, stimp, stoze, strop, sturch, sule, sump, sunge, surl, sutt, sym, sype, tace, tald, tark, tarmth, tarse, tatch, tays, teach, tearl, teigh, teign, telp, terb, terve, thak, thealt, thoar, thout, thrax, thwee, tidge, tiece, tinsie, tirl, tirt, titch, toal, toard, tob, tove, tord, torl, torse, toubt, traph, trave, treathe, treek, treen, trest, troes, trome, trool, troom, trouch, trunt, tuede, tulf, tunch, tusp, tymph, vaight, valf, vant, vaud, vect, veef, ver, vime, vinn, voe, voint, vor, vought, vour, vub, vyme, waith, weck, weem, weeth, weg, welse, wese, wesk, wict, wolt, wompt, worpse, wouge, wounge, woute, wug, wulch, wunk, wurve, yalt, yarce, yarm, yied, yife, yin, yince, yoops, yothe, yus, zale, zigh, zisle, zuct, zuss

From Sibley et al. (2006):

brandles, mccaftle, pruttor, schrins, goader, guyara, bocenie, hinensly, barts, slotlare, endaces, honter, proter, baplers, aterkan, stadio, movenla, nolin, skeldman, barss, lanter, netools, michlet, korlan, crerly, desslap, murses, slueiks, adline, vicorted, matiad, mante, oislen, mabacets, shucars, wongiers, cloin, pavile, letoras, gratete, adrins, ilblinit, claxtin, weellee, forple, arkisson, sermen, garley, starme, sirroned, sedaws, sirkhio, ravenies, meadeved, xorores, detrikk, erlar, irfamler, sartene, buratian, marmas, culglen, freedhad, rorry, funter, marrita, stoplans, baclater, strultry, payborty, marmers, sonetle, mazlits, wiske, marpsmar, waman, cleerery, branten, arasslam, basly, doclone, blymen, bactracs, bekor, dergoin, bardey, randra, mamber, torsen, sontalle, sunglers, bomod, andendun, mayuz, cantur, corcess, pegred, adarnid, parders, athire, bulli, mithakes, gerasey, morprine, undreide, reebold, krentifs, roadham, streans, monedea, restling, rotess, scharx, shoston, naintira, sarby, macee, crultie, busla, demutily, steddend, treakale, ganzes, deprelen, wanchort, voruranc, dertlee, bourties, strares, hielatle, scheoned, neluly, caldrer, crertle, sanild, runey, stedy, ronte, hamone, daney, muellen, namarm, axcises, bertol, paluera, arlens, dinbuese, hengters, amanding, slopheng, morarier, mause, deblan, purints, halaren, realens, reier, pather, allerte, brosed, borels, omich, licten, goitte, hobslon, sintars, ciner, bener, ellay, banborn, corleder, mayaste, prits, bresling, dleised, luehter, faricine, sriteanf, poded, hebrants, wathrole, overetle, aqbel, intail, tramon, barlo, garaley, brantade, plorm, marik, banfand, hallens, aliedo, mcrangy, wortover, canian,

frinired, ppoled, burreo, echoc, ansints, mertile, manher, outsmans, gertisos, eangs, cubar, wanstas, gonlan, senda, rones, wradley, panen, alele, slulth, goiffle, sysbel, ginge, maralia, ovaliade, sinos, derto, pasle, shibane, coinlels, sarser, migarle, bencham, roren, werolers, cuvinend, swynh, extin, marsla, mancups, sloners, cerel, reermoms, teflind, carafy, ryrinf, hesnane, laranto, sililony, mollo, shintey, sungy, rarik, cilts, leatlean, lanawa, casled, malton, hartate, sinsle, roured, reanener, clillen, beprars, moribins, acorale, poschan, gentarys, galler, wingins, dotwoen, pasered, traffles, chens, cabbe, mahtmas, camerews, sters, forewald, praiser, prare, mariacce, mymad, ambel, sarla, angolica, marri, abany, horled, astalle, parnage, ginoeing, pranthes, ingabic, edibbe, derfine, neles, jarts, schers, clags, boalans, medrert, bainne, bener, dechomal, wirlars, attonter, dedbush, crerley, baped, maynew, gaged, cerdie, veeping, reden, lutchter, meassin, katene, beane, manten, coosale, contiir, schunter, cossline, shainens, hattlas, kerer, shamital, mollon, radenla, derfer, ballics, corshwin, pachter, namacan, optlaurs, stoiled, plavan, haker, pooneer, delaser, dasten, kinted, surkett, benem, graire, beess, sluaggle, catiry, bronto, rerts, meneliny, teveled, serla, nefre, outroins, adelben, carke, scepen, alakara, winlo, liscare, deasasle, fritrobe, mabes, seued, tento, pransar, atterins, welkday, labermer, debsane, sussy, matloon, antan, tinenity, prenhtan, jurded, aumane, hamris, sluck, tunkec, chabes, clalifs, cagrand, miendie, cinawork, lerer, bcaser, slino, leror, mintoled, mapoen, anters, cestfeld, massikes, dongane, slarerce, indertan, lurlis, slaus, cticson, diapen, borled, shean, hieds, tirpins, shany, tonkies, bersky, sloner, marler, joithled, armeter, harren, battly, mawlon, benlen, rimelve, fibes, dupily, chastesh, battes, mandan, penket, leuthan, penscs, mislerl, gollen, deder, broins, ranekes, quins, stitento,

hasbong, tatzir, bourz, belloute, bollonts, mondels, eretiont, cears, anere, fellay, sharen, roilish, tramo, pountes, binler, maken, regies, spokel, vellecen, alenting, devanies, wleas, ackhart, hammary, runover, hantend, mecee, overn, dedeh, clited, masous, mozede, minering, senty, lours, ancoiker, satenens, weaks, teres, sleele, rorerel, admer, mimpun, intied, zerly, cring, nintley, cauteen, baitedi, majans, ricada, manzei, srontong, coens, lallaw, aluation, demeine, sylics, deasho, yothie, feporbar, moyed, ancomsen, sagarane, sawante, brinled, ravirels, zadela, bonhem, anery, runin, breaked, filsly, krintler, uplirler, munrey, lainhted, seets, boponts, loughlen, foonine, stipars, carabel, shratler, tarrias, scront, sleers, lantlans, feeddirg, radeau, pleaver, ratianed, worron, sherpins, enhers, paila, hoocs, pells, serate, inatetes, cadering, shugling, slickler, frirors, heoing, ediens, bathran, rankburs, caured, payman, binth, aished, bouble, husher, merceted, amarinan, aldege, camts, stonk, brerts, gatewaan, dinda, beins, slerts, lunion, altens, mitled, berse, loalen, flandarg, orachane, borestre, ronty, maute, banem, molus, hoinn, mants, moncinte, happisty, slandash, cades, slenz, oxtex, banue, dexcuges, roins, scins, seren, dexed, stexhale, plarters, dosarcan, shiriner, lotingke, ancobano, benan, wagen, indintly, slandera, anrovils, airie, labne, amlie, stors, tarlo, ourle, rorry, rolar, lired, ayten

WORK CITED

## WORK CITED

- Aderman, D., & Smith, E. E. Expectancy as a determinant of functional units in perceptual recognition. *Cognitive Psychology*, 1971, 2, 117-129.
- Andrews, S. (1989). Frequency and neighborhood effects on lexical access: Activation or search? *Journal of Experimental Psychology: Learning, Memory and Cognition*, 15, 802-814.
- Andrews, S. (1997). The effect of orthographic similarity on lexical retrieval: Resolving neighborhood conflicts. *Psychonomic Bulletin & Review*, 4, 439-461.
- Aslin, R. N., Saffran, J. R., & Newport, E. L. (1999). Statistical learning in linguistic and nonlinguistic domains. In B. MacWhinney (Ed.), *Emergence of language*. (pp. 359-380). Mahwah, NJ: Lawrence Erlbaum.
- Auer, E. T., Jr., & Luce, P. A. (2003). Probabilistic phonotactics in spoken word recognition. In C. T. McLennan, P. A. Luce, G. Mauner, & J. Charles-Luce (Eds.). *University at Buffalo working papers on language and perception* (Vol. 2, pp. 164-202). Buffalo, NY: University at Buffalo.
- Bailey, T.M., & Hahn, U. (2001). Determinants of wordlikeness: Phonotactic or lexical neighborhoods? *Journal of Memory and Language*, 44, 568-591.
- Balota, D. A. & Chumbley, J. I. (1984). Are lexical decisions a good measure of lexical access? The role of word frequency in the neglected decision stage. *Journal of Experimental Psychology: Human Perception & Performance*, 10, 340-357.
- Balota, A. A., Cortese, M. J., Hutchison, K. A., Neely, J. H., Nelson, D., Sompson, G. B., et al.(2002). The English lexicon project: A web-based repository of descriptive and

- behavioral measures for 40,481 English words and nonwords.  
<http://elixon.wustl.edu>:Washington University.
- Balota, D. A., Cortese, M. J., Sergent-Marshall, S. D., Spieler, D. H., & Yap, M. (2004). Visual word recognition of single-syllable words. *Journal of Experimental Psychology: General*, *133*, 283-316.
- Balota, D.A., & Spieler, D.H. (1998). The utility of item-level analyses in model evaluation: A reply to Seidenberg and Plaut. *Psychological Science*, *9*, 238-240.
- Barlow, H. (1972). Single units and sensation: A neuron doctrine for perceptual psychology. *Perception*, *1*, 371-394.
- Baron, J., & Thurston, I. (1973). An analysis of the word-superiority effect. *Cognitive Psychology*, *4*, 207-228.
- Behrmann, M., & Bub, D. (1992). Surface dyslexia and dysgraphia: Dual routes, single lexicon. *Cognitive Neuropsychology*, *9*(3), 209-251.
- Bergman, M. W., Hudson, P. T., & Eling, P. A. (1988). How simple complex words can be: Morphological processing and word representations. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology*, *40*, 41-72.
- Binder J.R., Frost J.A., Hammeke T.A., Cox R.W., Rao S.M., Prieto T. (1997). Human brain language areas identified by functional magnetic resonance imaging. *Journal of Neuroscience*; *17*: 353-362.
- Binder JR, Frost J.A., Hammeke T.A., Bellgowan P.S.F., Springer J.A., Kaufman J.N., Possing E.T. (2000). Human temporal lobe activation by speech and nonspeech sounds. *Cerebral Cortex*; *10*: 512-528.
- Bishop, C. M. (1995). Neural networks for pattern recognition. Oxford: Oxford University Press.

- Bonte, M. L., Mitterer, H., Zellagui, N., Poelmans, H., & Blomert, L. (2005). Auditory cortical tuning to statistical regularities in phonology. *Clinical Neurophysiology*, 116(12), 2765-2774.
- Breedin, S. D., Saffran, E. M., & Coslett, H. B. (1994). Reversal of the concreteness effect in a patient with semantic dementia. *Cognitive Neuropsychology*, 11, 617-660.
- Brown, R., & Hildum, D. C. (1956). Expectancy and the perception of syllables. *Language*, 32, 411-419.
- Bub, D., Cancelliere, A., & Kertesz, A. (1985). Whole-word and analytic translation of spelling to sound in a consemantic reader. In K. Patterson, J. C. Marshall & M. Coltheart (Eds.), *Surface dyslexia* (pp. 15-34). London: Erlbaum.
- Castles, A., & Coltheart, M. (1993). Varieties of developmental dyslexia. *Cognition*, 47, 149-180.
- Chun, M. M. (1997). Types and tokens in visual processing: A double dissociation between the attentional blink and repetition blindness. *Journal of Experimental Psychology: Human Perception and Performance*, 23(3), 738-755.
- Cohen, J. (1983). The cost of dichotomization. *Applied Psychological Measurement*, 7, 249-253.
- Cohen, J., Cohen, P., West, S. G., & Aiken, L. S. (2003). *Applied multiple regression/correlation analysis for the behavioral sciences* (3rd ed.). Mahwah, NJ, US: Lawrence Erlbaum Associates, Publishers.
- 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(4), 589-608.

- Coltheart, M., Patterson, K., & Marshall, J. C. (Eds.). (1980). *Deep dyslexia*. London: Routledge & Kegan Paul.
- Coltheart, M., Rastle, K., Perry, C., Langdon, R. & Ziegler, J. (2001). DRC: A dual route cascaded model of visual word recognition and reading aloud. *Psychological Review*, *108*, 204-256.
- Cortese, M. J, & Fugett, A. (2004). Imageability ratings for 3,000 monosyllabic words. *Behavior Research Methods, Instruments & Computers*, *36*, 384-387.
- Cutler, A. (1981). Making up materials is a confounded nuisance, or: Will we be able to run any psycholinguistic experiments at all in 1990? *Cognition*, *10*, 65-70.
- Cunningham, A. E., & Stanovich, K. E. (1991). Tracking the unique effects of print exposure in children: Associations with vocabulary, general knowledge and spelling. *Journal of Educational Psychology*, *83*, 264-274.
- Cutler, A., Mehler, J., Norris, D., & Segui, J. (1986). The syllable's differing role in thesegmentation of French and English. *Journal of Memory & Language*, *25*, 385-400.
- Dehaene, S., Le Clec'h, G., Poline, J., LeBihan, D., & Cohen, L. (2002) The visual word form area. A prelexical representation of visual words in the fusiform gyrus. *NeuroReport*, *13*:321-325.
- Dufva, M., Niemi, P. & Voeten, M. (2001). The role of phonological memory, decoding, and comprehension skills in reading development: From preschool to grade 2. *Reading and Writing: An Interdisciplinary Journal*, *14*, 91-117.
- Elman, J.L. (1990). Finding structure in time. *Cognitive Science*, *14*, 179-211.
- Farah, M. J., Stowe. R. M.. & Levinson, K. L. (1996). Phonological dyslexia: loss of a reading-specific component of the cognitive architecture? *Cognitive Neuropsychology*, *13*, 849,866.

- Fiez, J. A. & Petersen, S. E., (1998) Neuro-imaging studies of word reading. *Proceedings of the National Academy of Sciences of the USA*, 95, 814-921.
- Fodor, J., & Pylyshyn, Z. (1988). Connectionism and cognitive architecture: A critique. *Cognition*, 28, 3-71.
- Forster, K. I. & Chambers, S. M. (1973). Lexical access and naming time. *Journal of Verbal Learning and Verbal Behavior*, 12, 627-635.
- Friedman, R. B. (1996). Phonological text alexia: poor pseudoword reading plus difficulty reading functors and affixes in text. *Cognitive Neuropsychology*, 13, 869-886.
- Frisch, S. A., Large, N. R., Zawaydeh, B., & Pisoni, D. B. (2001). Emergent phonotactic generalizations in English and Arabic. In J. Bybee & P. Hopper (Eds.), *Frequency and the emergence of linguistic structure* (pp. 159-179). Amsterdam: John Benjamins.
- Garan, E.M. (2002). Beyond the smoke and mirrors: A critique of the National Reading Panel report on phonics. In R.L. Allington (Ed.), *Big brother and the national reading curriculum: How ideology trumped evidence* (pp. 90-111). Portsmouth, NH: Heinemann.
- George, D., & Mallery, P. (2003). *SPSS for Windows step by step: A simple guide and reference*. 11.0 update (4th ed.). Boston: Allyn & Bacon.
- Glushko, R.J. (1979). The organization and activation of orthographic knowledge in reading aloud. *Journal of Experimental Psychology: Human Perception and Performance*, 5, 674-691.
- Goswami, U., & Ziegler, J. C. (2006). A developmental perspective on the neural code for written words. *Trends in Cognitive Sciences*, 10, 142-143.

- Graham, K. S., Hodges, J. R., & Patterson, K. (1994). The relationship between comprehension and oral reading in progressive fluent aphasia. *Neuropsychologia*, 32(3), 299–316.
- Grainger, J., Bouttevin, S., Truc, C., Bastien, M., & Ziegler, J. (2003). Word superiority, pseudoword superiority, and learning to read: A comparison of dyslexic and normal readers. *Brain & Language*, 87, 432-440.
- Greicius, M. D., Krasnow, B., Reiss, A. L., & Menon, V. (2003). Functional connectivity in the resting brain: A network analysis of the default mode hypothesis. *Proceedings of the National Academy of Sciences of the USA*, 100, 253-258.
- Grossberg, S. (1980). How does a brain build a cognitive code? *Psychological Review*, 87, 1-51.
- Harm, M. W., & Seidenberg, M. S. (1999). Phonology, reading acquisition, and dyslexia: Insights from connectionist models. *Psychological Review*, 106, 491-528.
- Harm, M. W., & 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.
- Hickok, G. & Poeppel, D. (2007). The cortical organization of speech processing. *Nature Reviews Neuroscience*, 8, 393-402.
- Hockett, C.F. (1960). The origin of speech. *Scientific American*, 203, 88-96.
- Horowitz, B., Rumsey., J. M., & Donogue., B.C. (1998). Functional connectivity of the angular gyrus in normal reading and dyslexia. *Proceedings of the National Academy of Sciences of the USA*, 95, 8939-8944.
- Jared, D. (2002). Spelling-sound consistency and regularity effects in word naming. *Journal of Memory & Language*, 46, 723-750.

- Jones, M. N., & Mewhort, D. J. K. (2007). Representing word meaning and order information in a composite holographic lexicon. *Psychological Review*, *114*, 1-37.
- Jordan, M. I. (1986). Serial order: A parallel distributed processing approach (No. 8604 ICS Technical Report): University of California at San Diego, La Jolla, CA.
- Jurafsky, D. (2003). Probabilistic Modeling in Psycholinguistics: Linguistic Comprehension and Production. In R. Bod, J. Hay & S. Jannedy (Eds.), *Probabilistic Linguistics*. Cambridge, MA: MIT Press. 48
- Jusczyk, P. W., Friederici, A. D., Wessels, J. M., & Svenkerud, V. Y. (1993). Infants' sensitivity to the sound patterns of native language words. *Journal of Memory & Language*, *32*, 402-420.
- Jusczyk, P. W., Luce, P. A., & Charles-Luce, J. (1994). Infants' sensitivity to phonotactic patterns in the native language. *Journal of Memory & Language*, *33*, 630- 645.
- Kello, C. T. (2003). The emergence of a double dissociation in the modulation of a single control parameter in a nonlinear dynamical system. *Cortex*, *39(1)*, 132-134.
- Kello, C. T., & Plaut, D. C. (2003). Strategic control over rate of processing in word reading: A computational investigation of the tempo-naming task. *Journal of Memory and Language*, *48*, 207-232.
- Kello, C. T., Sibley, D. E., & Colombi, A. (2004). Using simple recurrent networks to learn fixed-length representations of variable-length strings. In Proceedings of the AAAI Symposium on Compositional Connectionism. Washington, DC.
- Kello, C. T., Sibley, D. E., & Plaut, D. C. (2005). Dissociations in performance on novel versus irregular items: Single-route demonstrations with input gain in localist and distributed models. *Cognitive Science*, *29*, 627-654.
- Kello, C. T. (2006). Considering the junction model of lexical processing. To appear in S. Andrews (Ed.), *All about words: Current research in lexical processing*. Sydney: Psychology Press.

- Knowlton, B.J., & Squire, L.R. (1993). The learning of categories: Parallel brain systems for item memory and category knowledge. *Science*, 262, 1747–1749.
- Levelt, W. J. M., & Wheeldon, L. (1994). Do speakers have access to a mental syllabary? *Cognition*, 50, 239-269.
- Marchand, y. & Damper, R. I. (2000) A multi-strategy approach to improving pronunciation by analogy, *Computational Linguistics* 26, pp. 195–219.
- Marcus, M., Santorini, B., & Marcinkiewicz, M. A. (1993). Building a large annotated corpus of English: The Penn Treebank. *Computational Linguistics*, 19, 313-330.
- Maris, E. (2002). The role of orthographic and phonological codes in the word and the pseudoword superiority effect: An analysis by means of multinomial processing tree models. *Journal of Experimental Psychology: Human Perception & Performance*, 28, 1409-1431.
- Marshall, J. C., & Newcombe, F. (1966). Syntactic and semantic errors in paralexia. *Neuropsychologia*, 4, 169–176.
- Marshall, J.C. & Newcombe, F. (1973). Patterns of paralexia: A psycholinguistic approach. *Journal of Psycholinguistic Research*, 2, 175-199.
- McCandliss, B. D., Cohen, L., & Dehaene, S. (2003). The visual word form area: Expertise for reading in the fusiform gyrus. *Trends in Cognitive Sciences*, 7(7), 293-299.
- 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, 375-405.
- McClelland, J.L., McNaughton, B.L. & O'Reilly, R.C. (1995). Why there are complementary learning systems in the hippocampus and neocortex: Insights from the

- successes and failures of connectionist models of learning and memory. *Psychological Review*, 102, 419-457.
- Miller, G. A. (1990). WordNet: An on-line lexical database. *International Journal of Lexicography*, 3, 235-312.
- Morton, J. (1970). A functional model for memory. In D. A. Norman (Ed.), *Models of human memory* (pp. 203-254). New York: Academic Press.
- Mummery, C. J., Ashburner, J., Scott, S. K., & Wise, R. J. (1999). Functional neuroimaging of speech perception in six normal and two aphasic subjects. *Journal of the Acoustic Society of America*, 106, 449-57.
- New, B., Ferrand, L., Pallier, C., & Brysbaert, M. (2006). Re-examining word length effects in visual word recognition: New evidence from the English Lexicon Project. *Psychonomic Bulletin & Review*, 13, 45-52.
- Okada, K. & Hickok, G. (2006). Identification of lexical-phonological networks in the superior temporal sulcus using fMRI. *Neuroreport*, 17, 1293-1296.
- Page, M. (2000). Connectionist modeling in psychology: a localist manifesto. *Behavioural and Brain Sciences*, 23(4):443-67.
- Patterson, K. E. (1982). The relation between reading and phonological coding: Further neuropsychological observations. In A.W. Ellis (Ed.), *Normality and pathology in cognitive functions* (pp. 77-111). San Diego, CA: Academic Press.
- Patterson, K., & Ralph, M. A. (1999). Selective disorders of reading? *Current Opinion in Neurobiology*, 9(2), 235-239.
- 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.

- Pinker, S. (1999) *Words and Rules: The Ingredients of Language*. New York: Basic Books.
- Pinker, S., & Prince, A. (1988). On language and connectionism: Analysis of a parallel distributed processing model of language acquisition. *Cognition*, 28, 73-193.
- Plaut, D. C. and Booth, J. R. (2000). Individual and developmental differences in semantic priming: Empirical and computational support for a single-mechanism account of lexical processing. *Psychological Review*, 107, 786-823.
- Plaut, D. C., Behrmann, M., Patterson, K., & McClelland, J. L. (1993, November). Impaired oral reading in surface dyslexia: Detailed comparison of a patient and a connectionist network [Abstract]. *Proceedings of the 34th Annual Meeting of the Psychonomic Society* (p. 48). Washington, DC.
- Plaut, D. C., & Shallice, T. (1993). Deep dyslexia: A case study of connectionist neuropsychology. *Cognitive Neuropsychology*, 10, 377-500.
- Plaut, D. C. (1997). Structure and function in the lexical system: Insights from distributed models of word reading and lexical decision. *Language and Cognitive Processes*, 12, 767-808.
- 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.
- Price, C. J., Howard, D., Patterson K., Warburton, E. A., Friston, K. J., & Wise., (1998). A functional neuroimaging description of two deep dyslexic patients. *Journal of Cognitive Neuroscience*, 10, 303-315.
- Rastle, K., & Coltheart, M. (2000). Lexical and nonlexical print-to-sound translation of disyllabic words and nonwords. *Journal of Memory & Language*, 42, 342-364.

- Raymer, A. M., & Berndt, R. S. (1994). Models of word reading: Evidence from Alzheimer's disease. *Brain and Language*, 47, 479-482.
- Reicher, G. M. (1969). Perceptual recognition as a function of meaningfulness of stimulus material. *Journal of Experimental Psychology*, 81, 275-280.
- Resnick, Lauren B. & Phyllis A. Weaver. (1976). Theory and Practice of Early Reading. Chall's 29-55.
- Rohde, D. L. T., Gonnerman, L. M., & Plaut, D. C. (submitted). An improved method for deriving word meaning from lexical co-occurrence. *Cognitive Science*.
- Rumelhart, D. E., McClelland, J. L., and the PDP Research Group. Parallel Distributed Processing: Explorations in the Microstructure of Cognition. Volume 1, MIT Press, 1986.
- Rumelhart, D. E., Durbin, R., Golden, R., & Chauvin, Y. (1995). Backpropagation: The basic theory. In Chauvin, Yves (Ed); Rumelhart, David E (Ed) (1995) Backpropagation: Theory, architectures, and applications Developments in connectionist theory (pp 1-34).
- Saffran, J. R. (2001). Words in a sea of sounds: The output of infant statistical learning. *Cognition*, 81, 149-169.
- Saffran, J.R., Aslin, R.N., & Newport, E.L. (1996). Statistical learning by 8-month old infants. *Science*, 274, 1926-1928.
- Schwartz, M. F., Marin, O. M., & Saffran, E. M. (1979). Dissociations of language function in dementia: A case study. *Brain and Language*, 7, 277-306.
- Sereno, J. A. (1986). Stress pattern differentiation of form class in English. *Journal of the Acoustical Society of America*, 79, S36.

- Seidenberg, M. S., Waters, G. S., Barnes, M. A., & Tanenhaus, M. K. (1984). When does irregular spelling or pronunciation influence word recognition? *Journal of Verbal Learning and Verbal Behavior*, *23*, 383-404.
- Seidenberg, M. S. and Plaut, D. C. (2006). Progress in understanding word reading: Data fitting versus theory building. in S. Andrews (Ed.), *From inkmarks to ideas: Current issues in lexical processing*. Psychology Press: Hove, UK.
- Seidenberg, M. S., Plaut, D. C., Petersen, A. S., McClelland, J. L., & McRae, K. (1994). Nonword pronunciation and models of word recognition. *Journal of Experimental Psychology: Human Perception and Performance*, *20*, 1177-1196.
- Seidenberg, M. & McClelland, J. (1989). A distributed developmental model of word recognition and naming. *Psychological Review*, *96*, 523-568.
- Seidenberg, M. S., & MacDonald, M. C. (1999). A probabilistic constraints approach to language acquisition and processing. *Cognitive Science*, *23*, 569-588.
- Shallice, T. (1988). *From neuropsychology to mental structure*. Cambridge, UK: Cambridge University Press.
- Shallice, T., & Warrington, E. K. (1980). Single and multiple component central dyslexic syndromes. In M. Coltheart, K. Patterson, & J. C. Marshall (Eds.), *Deep Dyslexia*, (pp. 119-145), London: Routledge & Kegan Paul.
- Sibley, D. E., Kello, C. T., Plaut, D. C., & Elman, J. L. (In Press). Learning representations of linguistic sequences. To be published in *Cognitive Science*.
- Sibley, D. E. & Kello, C. T. (2004). Computational explorations of double dissociations: Modes of processing instead of components of processing. *Cognitive Systems Research*, *6*, 61-69.

- Spieler, D. H., & Balota, D. A. (1997). Bringing computational models of word naming down to the item level. *Psychological Science*, 8, 411-416.
- Stone, G. O., Vanhoy, M., & Van Orden, G. C. (1997). Perception is a two-way street: Feedforward and feedback phonology in visual word recognition. *Journal of Memory & Language*, 36, 337-359.
- Toglia, M. P., & Battig, W. F. (1978). Handbook of semantic word norms. Hillsdale, NJ: Erlbaum.
- Vitevitch, M. S., & Luce, P. A. (1999). Probabilistic phonotactics and neighborhood activation in spoken word recognition. *Journal of Memory and Language*, 40, 374-408.
- Weekes, B. S. (1997). Differential effects of number of letters on word and nonword naming latency. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology*, 50A, 439-456.
- Wheeler, D. D. (1970). Processes in word recognition. *Cognitive Psychology*, 1, 59-86.
- Yap, M. (2007). Visual word recognition: Explorations of megastudies, multisyllabic words, and individual differences. PhD thesis, Washington University.
- Yarkoni, T., Balota, D. A., & Yap, M. J. (submitted). Levenshtein distance measures of orthographic and phonological similarity in visual word recognition.
- Ziegler, J. C., & Goswami, U. (2005). Reading acquisition, developmental dyslexia, and skilled reading across languages: a psycholinguistic grain size theory. *Psychological Bulletin*, 131, 3-29.
- Ziegler, J. C., Perry, C., Jacobs, A. M., & Braun, M. (2001). Identical words are read differently in different languages. *Psychological Science*, 12, 379-384.

## CURRICULUM VITAE

Daragh E. Sibley received a B.S. in Psychology and a minor in Chemistry, from Virginia Polytechnic and State University in Blacksburg, Va. He then entered the graduate program in George Mason University's Psychology department. In 2004, Daragh received an M.A. from the Human Factors and Applied Cognition program. This dissertation was submitted in 2008, in partial completion of his Ph.D. in Psychology. Daragh has accepted a post doctorate position and plans to pursue a career in academia.