

8-23-2016

An Investigation of Lexical Coherence in Novel Word Learning

Ashlee Shaw

University of Connecticut - Storrs, a.metshell@gmail.com

Follow this and additional works at: <https://opencommons.uconn.edu/dissertations>

Recommended Citation

Shaw, Ashlee, "An Investigation of Lexical Coherence in Novel Word Learning" (2016). *Doctoral Dissertations*. 1250.
<https://opencommons.uconn.edu/dissertations/1250>

An Investigation of Lexical Coherence in Novel Word Learning

Ashlee Shaw, PhD

University of Connecticut, 2016

The Lexical Quality Hypothesis suggests that the difficulties exhibited by poor readers cascade from deficient representations of phonological, semantic, and orthographic dimensions in lexical memory. This invites questions of what kinds of individual differences in cognitive abilities might lead to differences in lexical quality. In this dissertation, I used artificial lexicon learning studies and individual differences measures of language- and memory-related skills in an effort to understand how differences in component abilities assumed to be important for novel word learning might lead to differences in lexical quality.

I manipulated relationships between phonological, orthographic, and/or semantic features of the artificial lexicon items, such that the novel items themselves had differing levels of lexical quality. The first experiment focused solely on relationships between phonology and semantics; the second and third experiments focused on phonology and orthography. The final experiment combined all three lexical elements into a single word-learning study.

The results of these experiments serve to support the tenets of the Lexical Quality Hypothesis, and suggest that in addition to the linguistic skills explicitly named in the theory, paralinguistic skills may also serve as (consequential) measures of individual lexical quality.

An Investigation of Lexical Coherence in Novel Word Learning

Ashlee Shaw

B.S., Vanderbilt University, **2008**

A Dissertation

Submitted in Partial Fulfillment of the

Requirements for the Degree of

Doctor of Philosophy

at the

University of Connecticut

2016

APPROVAL PAGE

Doctor of Philosophy Dissertation

An Investigation of Lexical Coherence in Novel Word Learning

Presented by

Ashlee Shaw, B.S.

Major Advisor _____
James S. Magnuson

Associate Advisor _____
Jay Rueckl

Associate Advisor _____
Heather Bortfeld

University of Connecticut
2016

Acknowledgements

Firstly, I am grateful to my committee: Dr. Jim Magnuson, Dr. Jay Rueckl, Dr. Heather Bortfeld, and Dr. Ken Pugh. Without their support, assistance, and guidance, this research would not have been possible.

Additionally, I am grateful to Charmane Thurmand and the Multicultural Scholars Program (MSP) for its support— both financially and in mentorship. I believe that programs such as the MSP are critical in retaining underrepresented students within academia; thank you for helping guide us through these oft-uncharted waters.

I would also like to send heartfelt thanks to members of the PAC, both graduate students and undergraduate RAs, for their support over the years.

Finally, to my family, in all its forms: to the “UConn Family” for always being there, even when we’re not on the same continent; to my BodyWise Family for helping me keep my heart (and head) healthy; to my mother, for keeping me grounded; to my husband Reed, for being my partner in all ways; and finally, to Nutmeg, for being the best (and cutest) writing partner anyone could ask for.

Table of Contents

Chapter 1: Introduction	1
Chapter 2: Experiment 1.....	15
Chapter 3: Experiments 2a and 2b.....	36
Chapter 4: Experiment 3.....	57
Chapter 5: Individual Differences Comparisons, Experiments 1 and 3	70
Chapter 6: Experiment 4.....	88
Chapter 7: General Discussion	108

Chapter 1: Introduction

Perfetti and Hart's (2002) Lexical Quality Hypothesis (LQH) posits that most difficulties with reading comprehension can be linked causally to difficulties with the strength and richness of an individual's lexical knowledge. According to the LQH, a high quality lexical representation incorporates detailed orthographic, semantic, and phonological information, as well as information about how these elements interact (Perfetti, 2007); the stronger and more specific the information contained within a lexical representation is, the more efficiently that word can be accessed during reading. According to this hypothesis, less skilled readers possess weak or

unclear lexical representations that are not optimal for efficient access, which cascades to problems with comprehension (Perfetti, 2007; Perfetti & Hart, 2002).

As the LQH suggests, the act of reading is complex: minimally, in order to successfully interpret text, one must not only successfully engage the visual system, but also the systems involved in accessing phonology and meaning. Therefore, one would expect that this notion of lexical-quality-as-richness-of-representation (and the ensuing difficulties that may result from poor-quality representations) would extend to spoken language processing, as well. Accordingly, there are multiple places in the process where a breakdown could occur, resulting in reading difficulty.

Many theories address the nature of reading difficulty with an eye on a particular locus—naturally, in these cases, the focus is on the more extreme instances of reading difficulty (e.g., dyslexia). In this chapter, I will address the evidence for and against some of the most compelling theories of reading difficulty, with the goal of demonstrating that the LQH can accommodate each of them. Then, I will describe questions of word learning that are left unanswered by the LQH. Finally, I will introduce the steps I take to address those questions.

Weaknesses in Sensory Processing

Reading is a multisensory process; as such, a number of theories point to weaknesses in sensory information processing as the culprit behind reading difficulty. Renvall and Hari (2003) point to weaknesses in auditory processing as the locus of reading difficulty in dyslexics, suggesting that dyslexics' sensory memory stores may be larger than typical readers'. This larger sensory store increases the likelihood of interference of incoming rapid auditory information, and ultimately, weakens the cortical representations necessary for typical reading acquisition (Renvall & Hari, 2003). They argue that dyslexics' difficulty with phonological processing of

words is merely a manifestation of this more general auditory weakness, as adult dyslexics show diminished auditory mismatch negativity (MMN) Event-Related Potentials (ERPs) to rapidly presented auditory stimuli when compared to non-impaired controls (Nagarajan, Mahncke, Salz, Tallal, Roberts, & Merzenich, 1999; Renvall & Hari, 2003). Additionally, they found this effect to hold during magnetoencephalographic recordings—when presenting adult dyslexics with an auditory mismatch task involving infrequent changes in tonal pitch, Renvall and Hari (2003) found "markedly" weaker responses in the left auditory cortex when compared to non-impaired controls. It is worth noting, though, that while some studies have found this effect in the general auditory domain, others have only found the diminished MMN only for speech sounds (Schulte-Körne, Deimel, Bartling, & Renschmidt, 1998; Schulte-Körne, Deimel, Bartling, & Renschmidt, 2001), suggesting that this weakness in auditory processing is only found in a subset of dyslexics, and is perhaps not a source of reading difficulty—it may simply co-occur with reading difficulty.

Attentional Modulation

Other theories suggest that weakness lies not in the auditory or visual systems themselves, but in modulating attention between the two. Hartley and Moore's (2002) Processing Efficiency Hypothesis suggests that rather than a weakness in processing temporal data, the locus of [poor readers]' weakness is in extracting relevant visual and acoustic information from noisy input. Since all natural input is produced in variously noisy settings, weaknesses in focusing on incoming visual and auditory information could cascade into weaknesses in representations for that information. To that end, Facoetti, Lorusso, Cattaneo, Galli, and Molteni (2005) examined the focused multimodal attention (FMA) shifting ability in children with dyslexia/specific reading disability (SRD). They assert that by focusing on an item, its neural representation is

enhanced, and stronger neural representations of items leads to "faster reaction times, improved sensitivity... and reduced interactions with flanking (environmental) stimuli" (Facoetti et al., 2005). Conversely, poor attention to stimuli can lead to poor representation in many post-perceptual arenas ("short-term memory, perceptual decisions, and voluntary responses").

Facoetti et al. (2005) assert that "sluggish attentional shifting" from one perceptual item to another is the factor behind SRD, as they found dyslexic children's ability to shift visual attention from one item to another to be "sluggish and asymmetric" when compared to controls (Facoetti, Lorusso, Paganoni, Cattaneo, Galli, & Mascetti, 2003; Facoetti, Paganoni, Turatto, Marzola, & Mascetti, 2000; Facoetti, Turatto, Lorusso, & Mascetti, 2001). Similarly, Hari, Valta, and Uutella (1999) found that dyslexic adults had a longer attentional blink when compared to typically developing adults. Facoetti and colleagues also found that SRD children had slower cross-modal (visuospatial/auditory) attentional shifting compared to age-matched and reading ability-matched controls (2005). They contend that difficulties in shifting attention leads to difficulties in processing crucial visual or auditory information—which, in terms of reading ability, leads to degraded development of phonemic representations (Facoetti et al., 2005).

Magnocellular Theory: Poor Thalamic Development

While some theories of reading difficulty point to sensory processing systems, and others to attentional shifting ability, another seeks to combine the two by pointing to the hardware (that is, the cells themselves, and not just their output). The magnocellular theory of developmental dyslexia (Stein, 2001) asserts that the difficulties experienced by dyslexics are both visual and auditory in nature, and are due to abnormal development of the magnocellular systems of the lateral and medial geniculate nuclei of the thalamus, respectively. When compared to control brains, the neurons in the magnocellular layers of the lateral and medial geniculate nuclei of

dyslexic brains tended to be smaller, and less pruned (Galaburda & Livingstone, 1993; Galaburda, Menard, & Rosen, 1994; Livingstone, Rosen, Drislane, & Galaburda, 1991). Since development of these areas typically occurs within the fourth or fifth month of fetal development, weaknesses in the magnocellular layers of the thalamus cascades into a number of problems that manifest most clearly in reading ability (Stein, 2001).

For both the visual and auditory domains, Stein asserts that the malformations lead to diminished temporal sensitivity in the areas that receive mainly magnocellular input. In the visual domain, this manifests as reduced (compared to controls) dorsal visual stream activity in response to moving targets (Eden, VanMeter, Rumsey, Maisog, & Zeffiro, 1996). Additionally, Stein cites evidence from Lovegrove, Martin, Blackwood, and Badcock (1980), which showed that, compared to non-impaired controls, dyslexics had impaired contrast sensitivity to sinusoidal gratings, "particularly at low spatial and high temporal frequencies" (Stein, 2001). Notably, this weakness in performance did not hold for gratings with high spatial frequencies, which would activate the parvocellular visual system (Stein, 2001), suggesting that dyslexics are not just bad at all visual tasks—just tasks that stress the magnocellular system.

Stein goes on to admit that these weaknesses are slight, and not found for all dyslexics—and instead, suggests that motion sensitivity, rather than spectral contrast, is a better measure of magnocellular input to the visual system (2001). In fact, research using random dot kinetograms (RDK), show that poor readers have significantly more difficulty detecting coherent motion in an array of dots when compared to age- and IQ-matched controls (Cornelissen, Bradley, Fowler, & Stein, 1994; Cornelissen, Richardson, Mason, Fowler, & Stein, 1994; Talcott, Hansen, Elikem, & Stein, 2000; Talcott, Hansen, Willis-Owen, McKinnell, Richardson, & Stein, 1998). Smaller magnocellular layers could mean that the receptive fields served by these neurons are also

smaller—indeed, dyslexics were worse at the RDK task at much higher densities than controls, reflecting comparative loss of density information (Talcott et al., 2000).

But what does this have to do with reading difficulty? Visual magnocellular activity is important for visual attention, visual search, and eye movements, all of which are worse in dyslexics compared to controls (Everatt, 1999; Iles, Walsh, & Richardson, 2000; Stein & Walsh, 1997). Moreover, visual motion sensitivity is positively correlated with one's ability to spell irregular words, regardless of reading ability (Castles & Coltheart, 1993; Talcott, Witton, McClean, Hansen, Rees, Green, & Stein, 2000). Proponents of the magnocellular theory hold that this is because a reader must rely on the visual, instead of phonological, features of an irregular word to spell it correctly. For example, in the case of a pseudohomophone test (e.g., choosing whether *rane* or *rain* is the correct spelling), participants cannot rely on phonological information to correctly answer, and must instead focus on visual information (Olson, Wise, Conners, Rack, & Fulker, 1989; Talcott et al., 2000).

Much of the arguments for the magnocellular theory of developmental dyslexia are in the visual domain, though Stein argues that auditory ability is affected as well. While some argue that weakness in the auditory domain is the only weakness that dyslexics have (as addressed at the beginning of the chapter), Stein contends that "only about one third of dyslexics" have mainly phonological weaknesses, while another third suffer solely from visual/orthographic weaknesses, and the remaining third suffer from visual and auditory problems "in almost equal proportions" (2001). Much like how spectral contrast sensitivity can predict visual skill in dyslexics, so, too, can sinusoidal frequency and amplitude modulations predict phonological skill (McAnally & Stein, 1996; Menell, McAnally, & Stein, 1999; Stein & McAnally, 1996; Talcott et al., 2000; Talcott, Witton, McClean, Hansen, Rees, Green, & Stein, 1999; Witton, Richardson,

Griffiths, Rees, & Green, 1997; Witton, Talcott, Hansen, Richardson, Griffiths, Rees, Stein, & Green, 1998). Of note, this is only true when processing the kinds of shifts in frequency modulation present in natural speech—when detecting higher rates, dyslexics perform as well as controls (Moore, 1989).

Additionally, according to Stein (2001), frequency and amplitude modulation sensitivity correlates "strikingly (highly)" with nonword reading (Talcott et al., 2000; Talcott et al., 1999; Witton et al., 1998). This, too, taps into phonological ability: according to Talcott and colleagues, after controlling for orthographic (homophone spelling) ability, sensitivity to auditory frequency modulation accounted for "nearly 25% of residual variance" in phonological skill (nonword reading), for both skilled and poor readers (2000).

But while the magnocellular theory of developmental dyslexia covers a wide range of findings, it still does not cover all instances of reading difficulty. In a sample of 30 dyslexic adults, Amitay, Ben-Yehudah, Banai, and Ahissar (2002) found that only six showed "pure" magnocellular deficits; the remaining 24 showed impaired performance in visual and auditory tasks that did not tax the magnocellular pathways (that is, they did not test temporal processing abilities). Amitay and colleagues (2002) took this as evidence for the multimodal attentional shifting views of dyslexia. Furthermore, Facoetti et al. (2005) argue that the deficits found in dyslexics with magnocellular abnormalities associate only with those dyslexics whose deficits are phonological. Finally, others contend that the difficulties Stein and others attribute to magnocellular processing are instead rooted in noise exclusion: noting that many studies testing the magnocellular hypothesis also used noisy visual displays, Sperling, Lu, Manis, and Seidenberg (2005) showed dyslexic and non-dyslexic children low- and high-noise images that were meant to differentially stimulate the magno- and parvocellular pathways. They found that it

was presence of visual noise in the stimuli, not the visual pathway stimulated, that created a difference in performance between groups: while dyslexics had a significantly higher contrast discrimination threshold than nondyslexics in the presence of visual noise, there was no difference in contrast thresholds between dyslexic and non-dyslexic children, for either the magno- or parvocellular displays, in the absence of visual noise. Sperling and colleagues also replicated these effects of visual noise on motion perception in both good and poor reading adults and children, suggesting that the evidence previously thought to be caused by a magnocellular deficit are most likely just noise-exclusion difficulties (Sperling, Lu, Manis, & Seidenberg, 2006).

Phonological Deficit Hypothesis: Focus on behavior

Perhaps, then, more insight into the loci of reading difficulty could be gained by focusing less on the biological mechanisms, and more on the behavior itself. To do so, I will shift our attention away from theories seeking a physical locus, and back to behavior (back both in terms of our discussion, and chronologically). The phonological deficit hypothesis (Shankweiler & Crain, 1986) states that poor readers have underdeveloped ‘phonological awareness’—the understanding that Spoken Words are comprised of individual phonemes—compared to non-struggling readers. This weakness in phonological representations leads to reading difficulty. Indeed, proponents of the phonological deficit hypothesis have shown that dyslexic readers improved markedly after phonologically based intervention, reading at the same level as non-struggling readers (Shaywitz, Shaywitz, Blachman, Pugh, Fulbright, Skudlarski, Mencl, Constable, Holahan, Marchione, Fletcher, Lyon, & Gore, 2004). However, it is not clear that this approach captures the full range of difficulties that exist in some poor readers.

For example, those with Specific Reading Comprehension Deficits (S-RCD) have intact phonological or word-level abilities (e.g., recognizing or decoding words) but struggle with reading comprehension (Cutting, Clements-Stephens, Pugh, Burns, Cao, Pekar, Davis, & Rimrodt, 2013). In contrast with dyslexic readers, those with S-RCD have intact representations of orthographic and phonological properties of lexical items, while their representations of lexical-semantic representations of items may be less secure. While less common than phonological dyslexia, S-RCD still affects anywhere from 3-10% of school-aged children (Cutting et al., 2013).

The Lexical Quality Hypothesis

This brief literature review shows that while deficit-specific theories of reading difficulty may account for some readers, no one deficit accounts for all readers. If the goal is to account for the entire set of readers, then perhaps one should consider the entirety of the mental lexicon—and the entirety of readers, including those who may not meet diagnostic criteria for disordered reading. Perfetti and Hart's (2002) Lexical Quality Hypothesis (LQH) also addresses the nature of reading difficulty; they accommodate the previously mentioned theories under the umbrella of the mental lexicon, stating that poor readers have impoverished (fuzzy, imprecise) representations of particular dimensions (phonology, semantics) or even specific items (words that are poorly learned and/or are atypical in dimensions like phonology or semantics) in their mental lexicons, and it is these poor representations that lead to poor reading comprehension. Importantly, the LQH is multidimensional; it posits that weakness in any of several dimensions (or multiple dimensions) of lexical representations (orthographic, phonological, semantic, or grammatical) could impact multiple dimensions and underlie an individual's difficulties with

reading. Additionally, Perfetti (2007) discusses how interactions among these dimensions may be another important factor (whether or not these elements are “tightly bound”).

The LQH is quite accommodating in its explanations of reading ability and difficulty. The ‘quality’ of any given lexical item can differ between individuals (one person may have more experience with either a single item, or have generally stronger or larger phonological abilities or lexical entries than another person; this holds for all readers, from the range of poor to superior), or it can differ among items, within an individual, for either one factor or many. For example, one lexical item may be of inherently higher quality than another: a high frequency item would have better quality than a low frequency item, as its features would be better reinforced through experience. Additionally, with frequency held constant, homophones, homographs, and homologues (which each have many-to-one mappings between their sounds, spellings, and/or meanings, respectively) all have inherently lower quality than those items with less ambiguity in the mapping between spelling, phonology, and meaning, as the overlap in lexical entries can be the basis for confusion (Perfetti (2007) calls this relationship between the features of a lexical entry “constituent binding”).

For example, take the word "*record*." It is both a homophone (e.g., "The stenographer took a *record* of the events" versus "It was no surprise that 'Get Lucky' won *Record* of the Year") and homograph (e.g., "I brought a camera to *record* the recital"). Here, one letter string is attached to three different concepts, and so likely to three different lexical entries. Thus, encounters with this letter string would lead to slower, more decision-laden processing than with a letter string that is only associated with one lexical item (given equal frequency, etc.). A notable strength, then, of the LQH is that it provides a basis for characterizing the heterogeneity

of strengths and weaknesses between readers and within a single reader's lexicon, rather than a “one size fits all” hypothesis.

However, it is worth mentioning there is a tension between the tenets of constituent binding in the LQH and the study of concepts and categories more generally. While constituent binding is an intuitive idea— of course, a good, complete lexical entry must contain many elements that are each well-learned and interact well with one another – it is not particularly well-defined in the LQH literature. Appealing to the related notion of *coherent covariation* in the categories and concepts literature may provide new insights. The theory-theory of semantic cognition suggests that the most ideal exemplars of a category are generated when they are consistent with a causal theory that provides the basis for the concept (Murphy & Medin, 1985). This contrasts with concept theories that suggest that categories are formed from a set of median features (e.g., prototype theory; Rosch, 1975), or from comparison to previous exemplars of a category (e.g., exemplar theory; Medin & Schaffer, 1978). McClelland and Rogers (2003) showed that each of these theoretically-defined categorization behaviors can emerge from simple statistical learning mechanisms, sans any explicitly-stated rules. Using a feed-forward, parallel distributed processing network, they illustrated how learning the features of items within a model can lead to accurate categorization while capturing key phenomena that motivated prototype and exemplar theories, as well as generalization consistent with the theory-theory. For example, through many epochs of exposure to the features of various birds (a bird that can fly, sing, etc.), the network was able to more quickly learn that a new bird, sparrow, could also fly, sing, etc., as the model was able to generalize from previously learned item-feature pairings (McClelland & Rogers, 2003).

In many ways, lexical quality is defined by typicality, and typicality is determined by exposure to linguistic input. The gradual, repeated exposure and correction utilized by McClelland and Rogers' (2003) network is not unlike the idea of binding in the LQH. The repeated exposure and corrections used in McClelland and Rogers' model is not a far cry from the repeated exposures and corrections (such as between a certain orthographic string and its pronunciation) typical of reading instruction. And just as the degree of constituent binding can vary (as in homophones, homologues, and homographs), so can the idea of coherence vary (canaries and sparrows are birds that fly; penguins are birds that do not). By appealing to the idea of 'coherent covariation' in the concept literature, we may be able to gain some insight into just what constituent binding entails—and how the typicality or recurrence of a certain lexical feature does (or does not) interact with the overall quality of a lexical entry.

However, there is tension between the potential effects of typicality and overlap. The more typical a word is—the more common its phonotactics, orthotactics, and semantics—the higher quality it should be, as repeated exposure to these common constituents of such a word would lead to stronger binding. But the more typical a word is (in phonology, orthography, and semantics), the more overlap it must have with other words. This dovetails with gang effects (activation boosts due to recurrent connections) thought to underlie the generally facilitatory effects of larger neighborhoods for written words (Rumelhart & McClelland, 1981), but is at odds with inhibitory effects of phonological neighborhood in the auditory domain (Luce & Pisoni, 1998) thought to follow from inhibition and the time course of the speech signal (McClelland & Elman, 1986). This leads to an intriguing question: are homophones an anomalous case (with full overlap, possibly independent of typicality), or might there be a “tipping point” of diminishing return where too high a level of typicality begins to have negative

effects? An answer suggested by the literature on concepts and categories is a resounding “no”; the more typical the feature sets are for an instance of a category, the better participants judge it to be a good exemplar of that category, and the more quickly they can process it (e.g., Rosch & Mervis, 1975). However, there is a crucial difference between natural categories and words, and the implications of typicality. When we talk about typicality for natural concepts, we are talking about categories with many members (BIRD has robin, sparrow, penguin, emu, etc.), and the scope of typicality is the set of features an instance has relative to the mean and/or structured distribution of features over all category members (Smith & Medin, 1981). Words are quite different. On the one hand, each word might best be considered its own category, but on the other, the scope of typicality is extremely coarse—essentially, if we were to try to calculate the prototype relevant for thinking about the LQH, it would be the (structured) mean of all words, encompassing all parts of speech (and not simply semantic-type features).

The research outlined in this dissertation will address a number of questions that explore the notion of lexical quality. Regarding constituent binding in the LQH, is one kind of binding (for example, phonological-semantic binding) easier or more difficult to learn than others? Additionally, when learning novel lexical items, can participants pick up on coherent covariation implicit in naming relationships? How do individual differences in reading and memory (and therefore, individual differences in lexical quality) influence learning rates or preferences? Finally, how can we resolve the tension between the concept literature and the notion of constituent binding?

In the following chapters, I will describe four experiments that explore these questions using artificial lexicon learning tasks; while they will not explicitly involve reading, using such tasks will allow for some control over the lexical quality of stimuli. Focus will initially be on

learning novel objects paired with novel, phonologically presented names, to begin with a simple form of learning. Two follow-up experiments will extend this paradigm by replacing novel objects with a novel orthography. The final experiment, also using novel stimuli, will use novel phonology, orthography, and semantic categories and features. Each experiment will manipulate the degree of constituent binding between visual objects and phonology and/or orthography. In the first four experiments, items will either be perfectly well-bound, or perfectly poorly-bound, between a visual stimulus and its name. In the final experiment, items' features will vary between categories: novel items' features can range from being perfectly well-bound on all levels for a category to perfectly poorly-bound on all levels; some items will be well-bound for a certain lexical dimension (e.g., phonology) but atypical on a different lexical dimension (e.g., semantics). Finally, I will examine how individuals' performance on learning and memory tasks predict performance in each of these learning tasks. In separating the components of words, I will be able to gauge how skilled readers' individual linguistic and memory abilities may impact how they learn the names of novel items, and relate these findings back to the LQH and other existing theories of reading ability and the mental lexicon.

Chapter 2: Experiment 1

Perfetti and Hart's (2002) Lexical Quality Hypothesis (LQH) posits that most difficulties with reading comprehension can be linked causally to difficulties with the strength and richness of an individual's word-level knowledge. A high quality lexical representation incorporates detailed orthographic, semantic, and phonological information; the stronger and more specific the information contained within a lexical representation is, the more efficiently that word can be accessed during reading. According to the LQH, less skilled readers possess weak or unclear lexical representations that are not optimal for efficient access (Perfetti & Hart, 2002). These impoverished lexical representations lead to problems at higher levels of language processing—such as with reading comprehension—though the exact mechanisms behind such a cascade of difficulty is unspecified (Perfetti & Hart, 2002).

The "triangle model" of visual word recognition (e.g., Harm & Seidenberg, 1999, 2004; Plaut, McClelland, Seidenberg, & Patterson, 1996; Seidenberg & McClelland, 1989) provides a mechanistic analog to the LQH that can illuminate trade-offs that impact typicality and coherence. The "triangle" refers to representational units for phonology, semantics, and orthography (Figure 1). The bidirectional phonology-semantics mappings are learned first, with later learning of orthography (phonology-orthography and orthography-semantic mappings). Activation passes through all parts of the network. Once orthography is brought on-line, when the network's task is to activate phonology given orthography, two pathways can contribute: orthography-to-phonology, of course, but also orthography-to-semantics-to-phonology. When the model is trained on English, regular patterns dominate the orthography-phonology connections (because they occur in more words, and so occur more frequently on average), while irregular spelling-sound patterns are more likely to involve semantic circuits (since the orthography-phonology connections are dominated by regular mappings). However, very frequent words with irregular orthography-phonological patterns, by virtue of their many (and possibly early) training opportunities, are more likely to integrate into the regular-dominated orthography-phonology connections. Thus, the more typical a spelling pattern is and the more it conforms to regularities (i.e., has greater coherence between its orthographic and phonological patterns), the more quickly it is learned and the more robustly it is activated. Irregular patterns will tend to require greater integration of semantics to achieve robust activation (in the context of the triangle model, they will depend more on orthographic-semantic-phonology connections), though frequency of occurrence mitigates this tendency. To the degree that phonological or semantic representations or phonology-semantic pathways are noisy or weak prior to

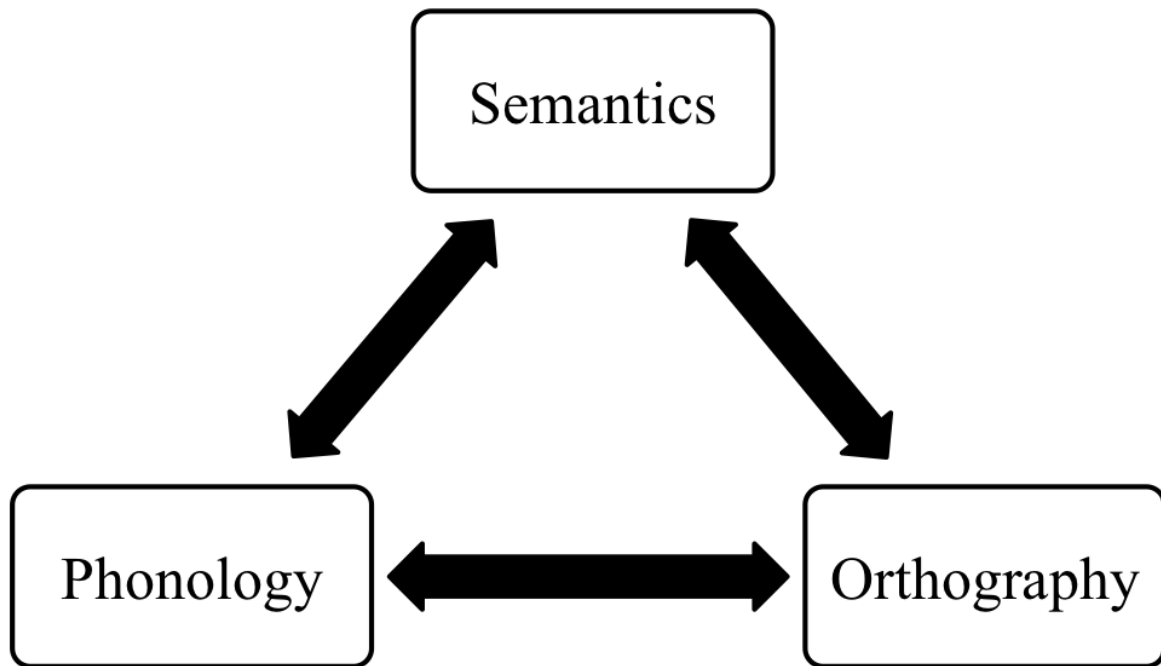


Figure 1. Simple Schematic of the Triangle Model.

orthographic learning, the model will be at a disadvantage when orthographic training begins (Plaut et al., 1996) – again, with greater risk for low-regularity, low-frequency lexical items.

Returning to the LQH, the mechanism embodied in the triangle model begs the question: what kinds of individual differences might lead to noisy or weak representations or pathways, and hence to low lexical quality? These could include sparse input/inexperience with a certain word or concept. However, in the absence of an actual language organ in the brain (as opposed to networks distributed over sensory and memory areas, as well as areas that seem specialized for aspects of language function; Cohen, Jobert, Le Bihan, & Dehaene, 2004; Cohen, Lehericy, Chochon, Lemer, Rivaud, & Dehaene, 2002; Hickock & Poeppel, 2004), other factors might contribute, such as sensory acuity, memory ability, associative learning ability, or the ability to

map information across modalities, such as from visual objects to names, or perhaps specific deficits in mapping phonology to print.

The researcher wishing to design experiments to disentangle these various potential contributions to language ability and lexical quality faces a quandary: individuals will arrive at the laboratory with a wide range of ability and experience. One way to examine acquisition of phonological-to-semantic connections while minimizing those differences in background is by using a spoken artificial lexicon (Magnuson, Tanenhaus, Aslin & Dahan, 2003). Utilizing an artificial lexicon allows us to tightly control properties of linguistic and visual materials and ensure that each participant has no prior experience with the stimuli, minimizing potential differences in lexical dimensions (e.g., word frequency) and preexisting semantic associations. Of course, this paradigm cannot neutralize all pre-experimental differences, but it allows us to observe any learning effects from the very beginning of the experiment. Thus, this paradigm allows us to put readers who vary in reading ability on a maximally similar level with regard to prior knowledge and language experience with our experimental items.

Hart and Perfetti (2008) used an artificial lexicon to examine the emergence of lexical interference and recovery for homophones during a semantic judgment task. In their experiment, Hart and Perfetti (2008) found that as participants had increased exposure and learning of novel homophones, high-frequency homophones reliably interfered with the activation of low-frequency counterparts. This was not the case at the start of training, when exposure to homophones was held constant (that is, there was no frequency difference between the two items in a homophone pair). Therefore, according to Hart and Perfetti (2008), not all homophones are created equal: greater exposure to one item from a homophone pair increases the "coherence"

amongst its features, leading to ease of lexical activation—and therefore, increased lexical interference upon exposure to a less-frequent homophone.

Artificial lexicon paradigms have been applied to the study of individual differences across a wide range of reading skill. Magnuson, Kukona, Braze, Johns, Van Dyke, Tabor, Mencl, Pugh, and Shankweiler (2010) found that performance on standard assessments like rapid auditory naming predicted the degree to which low-literacy adults exhibit lexical competition effects and how sensitive they are to coarticulation. However, while that project included dozens of language measures, it included only a few standardized assessments of non-linguistic abilities. This allows for several follow-up questions. Firstly, what sorts of individual differences might we observe in linguistic and non-linguistic abilities in a typical college sample (rather than the low-literacy adults from Magnuson et al., 2010)? Will those differences be compatible with the premises of the Lexical Quality Hypothesis (that is, will participants on the low end of linguistic ability—those most likely to have poor lexical representations, or poorer resources with which to create lexical representations—similarly lag behind their peers in learning novel words)? Alternatively, or perhaps in addition, might performance in learning new words be more strongly associated with simple learning (recognition memory) across domains (faces, objects, Spoken Words)?

We began our line of questioning by exploring the relationship between semantics and phonology. In Experiment 1, we examined whether performance scores on standardized tests of language ability or visual and language-related memory tasks could predict readers' ability to link new words to concrete visual objects. Our initial aim was to isolate levels of coherence among newly learned items, rather than on the interaction between coherence and typicality; as such, novel items in the experiment were either maximally or minimally coherent in their sound-

feature mappings. From the basis of the Lexical Quality Hypothesis, we predicted that language ability should be closely related to artificial lexicon learning: those who scored better on our language-related tasks should learn novel lexical items faster/have higher accuracy scores than those with lower scores on our language-related tasks. Our design also allows us to ask whether such differences are specific to language, or might apply more generally across domains (in this case, face and/or object recognition memory).

Methods

Participants

Seventy-six University of Connecticut undergraduates were participants in the experiment. All participants were native, monolingual English speakers who reported normal hearing and normal or corrected-to-normal vision.

Apparatus and Materials

Assessments. Five assessments were used to measure individuals' abilities in linguistic and nonlinguistic domains. Given the amount of time required to complete the artificial lexicon task, we endeavored to come up with the smallest number of tasks that tap into these domains. The tasks we used included a test of verbal working memory and syntactic ability (the Reading Span Task (RS; Daneman and Carpenter, 1980; van den Noort, Bosch, Haverkort, & Hugdahl, 2008) as well as tests of word and pseudoword reading efficiency (Test of Word Reading Efficiency [TOWRE], Torgeson, Wagner & Rashotte, 1997) that correlate highly with measures of reading fluency (Sabatini, Sawaki, Shore, & Scarborough, 2010). We also administered face, object, and Spoken Word recognition (old/new) tasks of our own construction, as measures of simple, non-linguistic memory and learning. The Spoken Word and object recognition tasks were meant to serve as a means of isolating participants' abilities in these arenas, respectively; face

recognition, while not explicitly an aspect of the experimental task, was intended to serve as a measure of visual recognition for a category of items participants have had years of experience with (as they have had with the English writing system).

The first assessment was the Reading Span Task (RS; Daneman and Carpenter, 1980; van den Noort et al, 2008), which is a measure of verbal working memory. In the RS, participants read multiple sets of 2-6 sentences aloud, with sentence lengths of 13-16 words; each sentence ends in a different word. After each group of sentences, participants are asked to recall the final words of each sentence in the group. Participants were tested on a total of 60 sentences (see van den Noort et al., 2008, for details).

The second assessment administered was the Test of Word Reading Efficiency (TOWRE; Torgeson, Wagner & Rashotte, 1997), which tested participants' word-level reading skills. This timed measure, normed for participants up to 24 years of age, quickly assesses the speed and accuracy of decoding and word recognition. It consists of two subtests. In the Sight Word Efficiency (SWE) subtest, the participant is presented with a list of printed real words and instructed to read aloud as many as possible in 45 seconds. Words in this subtest are arranged in order of decreasing frequency and increasing length. In the Phonetic Decoding Efficiency (PDE) subtest, the participant is presented with a list of pronounceable pseudowords and asked to decode aloud as many as possible in 45 seconds. The pseudowords in this list represent a variety of grapheme-phoneme correspondences and increase in difficulty as the test progresses. Thus, both subtests are designed to increase in difficulty while taxing the participant with added time pressure. Further, from the point of view of the Lexical Quality Hypothesis, the SWE subtest of the TOWRE should shed light on participants' ability to quickly access pre-existing lexical representations.

For the face recognition task, the stimuli were 50 faces taken from Nestor and Tarr (2008), which were approximately balanced in terms of gender and race: half of the faces presented were male, and the other female; there were approximately equal numbers of the self-reported races presented in the stimulus set (Black, Asian, Hispanic, White, and Mixed-race, Figure 2). During the exposure phase, participants were shown 25 faces for duration of 300 ms each, with a 300 ms inter-stimulus interval. During testing, participants were shown a total of 50 faces, and pressed a key to indicate whether she or he saw the face during exposure. Twenty-four of the faces (12 old, 12 new) during testing were presented in an alternate orientation (i.e., with a left- or right-facing profile rotated 30, 45, or 60 degrees).

The object recognition task included 150 realistically-rendered images of objects from the Tarr Object Databank (Figure 3, Images courtesy of Michael J. Tarr, Carnegie Mellon University, <http://www.tarmlab.org>). Objects were selected from 12 rough taxonomic categories (4-28 from each), and were judged by the experimenters to be roughly similar in visual salience. During the exposure phase, participants were shown 75 objects for a duration of 300 ms each, with a 300 ms inter-stimulus interval. During testing, participants were shown a total of 150 images, and pressed a key to indicate whether s/he recalled the object from the exposure phase. Half of the objects during testing were presented in an “alternate” orientation (rotated 90-180 degrees). Thirty-eight of the alternate orientations were of old objects, and 37 were new.

Finally, the old/new Spoken Word recognition task was constructed as follows: a total of 152 Spoken Words were recorded by two female speakers (words were 1-7 syllables; average syllables=2.4). Each of the speakers spoke half of the items for both the old and new sets (76 items total per speaker). Old items were categorized into a “same” or “different” condition —i.e., the speaker during the exposure phase either was or was not the same speaker during recognition



Figure 2. Examples of stimuli used in face recognition task. Stimuli were taken from Nestor and Tarr, 2008.



Figure 3. Examples of images used in the object recognition task, taken from the Tarr Object Databank (courtesy of Michael J. Tarr, Carnegie Mellon University, <http://www.tarrlab.org>).

testing. The instructions made clear that a word should be considered "old" even if the voice was not the same. During exposure, participants listened to 76 Spoken Words (300 ms inter-stimulus interval). During testing, participants heard 152 words and were instructed to press a key indicating whether the Spoken Word was heard during exposure or not.

Artificial lexicon experiment. The primary task was to learn the names of nine mushrooms. The mushrooms varied in two visual dimensions: they had one of three caps and one of three stems. Each mushroom had a two-syllable name, such as /pile/ ("pea-lay"). The names were combinations of three possible first syllables (/pi/ ("pea"), /do/ ("dough"), /gu/ ("goo")) and three possible second syllables (/le/ ("lay"), /va/ ("vah"), /sae/ (as in "sat")). The relationship between visual and phonological features was manipulated between participants. For participants in the "Correlated Naming" Condition (n=41), the syllables mapped directly onto visual properties of the mushrooms, such that the first syllable named the cap and the second named the stem (thus, the name of any mushroom with a particular cap would begin with the same syllable, and the name of any mushroom with a particular stem would have the same second syllable, Figure 4). In the "Uncorrelated Naming" Condition (n=35), visual and phonological features were completely uncorrelated, such that mushrooms with the same cap or stem had no phonological overlap in that dimension, and mushrooms with the same first or second syllable had no visual overlap in that dimension (see Figure 5).

Procedure

Phonological stimuli were recorded in a sound-attenuated booth, with a sampling rate of 44.1 kHz. They were then edited and checked for artifacts and clipping using Praat software. Visual stimuli were created by splicing together drawings of natural mushrooms from the DeAgostini Picture Library using Photoshop software, creating 9 novel mushrooms.
















	 (& #	 ' % #	 \$ % & #
 ! ' #	 ! ' (& #	 ! " % #	 ! " \$ % & #
 +, #	 +, (& #	 +, ' % #	 +, \$ % & #
) * #	) * (& #	) * ' % #	) * \$ % & #

Figure 4. Sample sound-picture pairings for stimuli in the Correlated Condition in Experiment 1. Note that actual first- and second-syllable labels were counterbalanced across participants, so that one participant's “/pile/” may have been another's “/gusae/,” but all relationships between mushroom caps and stems remained constant.

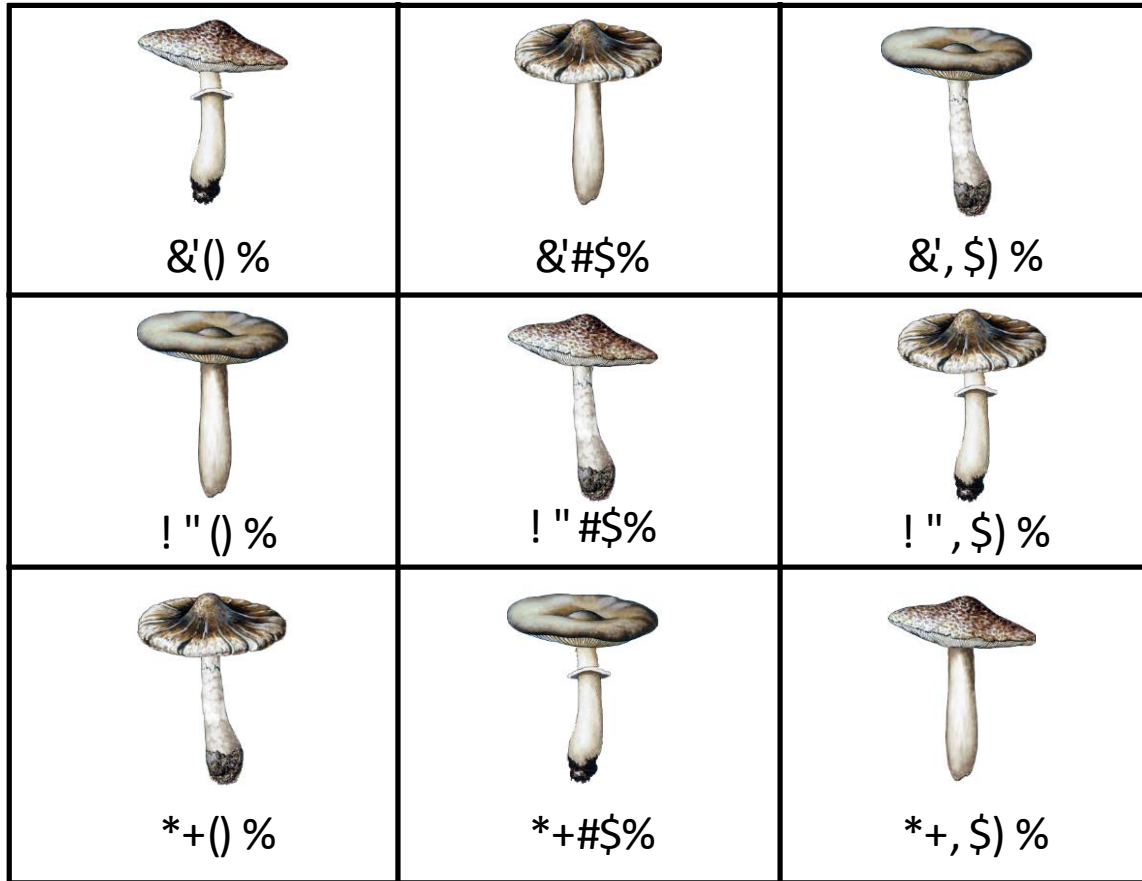


Figure 5. Sample sound-picture pairings for stimuli in the Uncorrelated Condition in Experiment 1. Note that actual first- and second-syllable labels were counterbalanced across participants, so that one participant's “/gule/” may have been another's “/pisae/”, but all relationships between the mushroom caps and stems remained constant.

The testing session began with the assessments and the exposure phases of the old/new tasks. These were followed by the artificial lexicon experiment, and then the test phases of the old/new tasks.

Participants were assigned randomly to the Naming Conditions. Participants were not informed about possible relationships between the names and visual features of the materials in either Condition. They were simply told they would learn the names of the objects, in a 2-alternative forced choice task. Each trial began when the participant clicked on a cross in the center of the screen; at the initial mouse click, reaction time recording was initiated. Immediately after the mouse click, the Target and distractor items joined the cross on the screen. Simultaneously, phonological stimuli were presented auditorally (via Sennheiser HD-595 headphones), in the form of instructions such as "Find /pile/". Participants responded by clicking on one of the mushrooms. If they clicked on the incorrect item, they heard an instruction to "try again." When the participant clicked on the correct item, the incorrect item disappeared, and they heard feedback like "That's right, that's /pile/," and then the trial ended (the entire feedback period was 2300 ms after clicking on the Target). A 1000 ms blank-screen inter-trial interval followed; to begin the next trial, participants clicked on a cross in the center of the computer screen. Every 24 trials, a progress report was displayed on the screen, telling the participant his/her percentage correct over the preceding 24 trials, and offering them an opportunity to take a break. Experimental Blocks consisted of 72 trials; over the course of a Block, participants were tested on each possible stimulus pairing. Trial order was pseudo-randomized in each Block so that each stimulus type was distributed equally over the Block. There were 5 Blocks, for a total of 360 trials. Stimuli were presented, and mouse responses recorded, using E-Prime software.

Results

In this chapter, experiment results will focus on accuracy and RT results within this experiment; individual differences comparisons between the phonological-visual item and phonological-orthographic experiments will be discussed in Chapter 5. To preserve a normal distribution of scores, accuracy scores were analyzed under a logit transformation, while reaction time scores were analyzed under a logarithmic transformation.

Accuracy and Reaction Time

Figures 6-11 illustrate average accuracy and RT scores over the course of the experiment. Both Naming Conditions showed an increase in accuracy (and corresponding decrease in RT) across each experiment and in each Condition.

Analyses of variance. Analyses of variance of Naming Condition on accuracy showed a main effect of Naming Condition on accuracy scores (Mean for Correlated = 0.93; Uncorrelated Mean = 0.71; $F(1, 74) = 139.8, p < 0.001$), as well as a main effect of Block [$F(4, 300) = 158.3, p < 0.001$], corroborating the visual inspection of accuracy data. Post hoc comparisons (using Tukey HSD contrasts) indicated significant accuracy increases ($p < 0.001$) from Block 1 to all following Blocks (Blocks 2, 3, 4, and 5), and Block 2 to all following Blocks (Blocks 3, 4, and 5). There was also a significant interaction between Naming Condition and Block [$F(4, 296) = 15.41, p < 0.001$]. Follow-up ANOVAs to unpack this interaction showed that, when separated by Naming Condition, the effect of Block on accuracy scores was still significant for each Naming Condition [Correlated Naming Condition: $F(4, 160) = 145.8, p < 0.001$; Uncorrelated Naming Condition: $F(4, 136) = 65.25, p < 0.001$]. However, post-hoc Tukey contrasts showed that the effect of Block on accuracy scores differed once separated by Naming Condition. In the Correlated Naming Condition, as when collapsed across Naming Conditions, there were

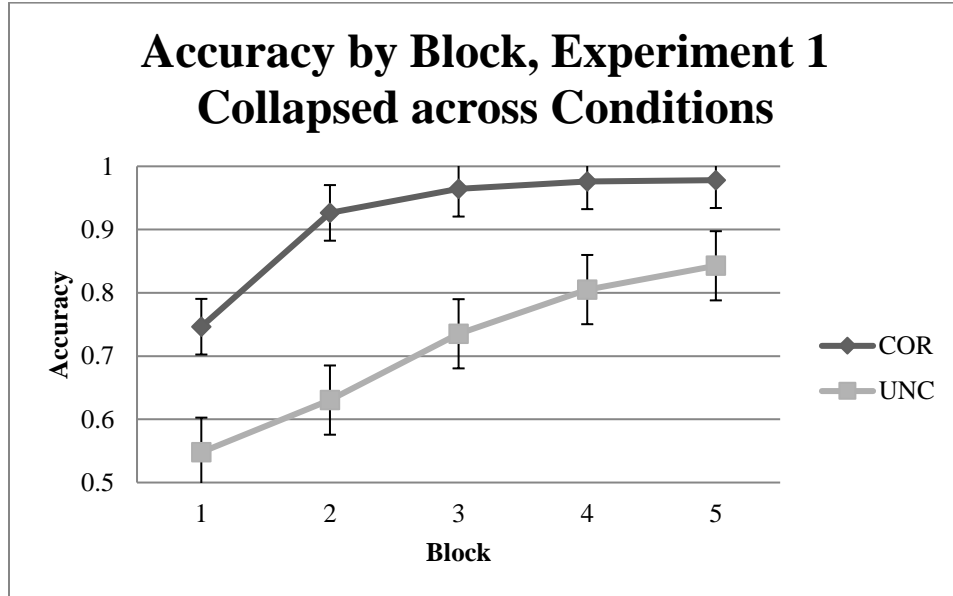


Figure 6. Accuracy scores for Experiment 1. Note: COR= Correlated Naming Condition; UNC= Uncorrelated Naming Condition.

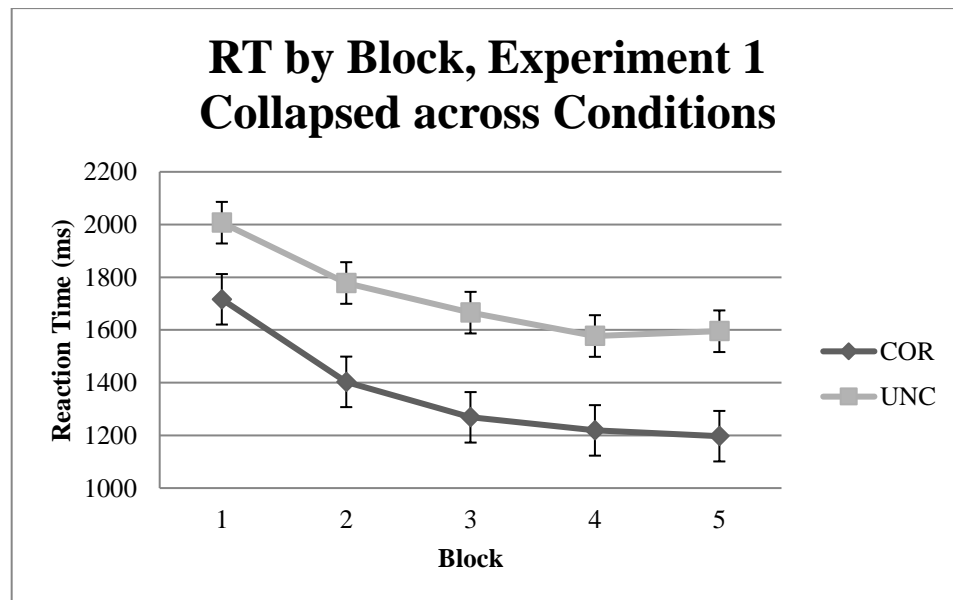


Figure 7. Reaction time for Experiment 1. Note: COR= Correlated Naming Condition; UNC= Uncorrelated Naming Condition.

significant accuracy increases ($p < 0.001$) from Block 1 to all following Blocks (Blocks 2, 3, 4, and 5), and Block 2 to all following Blocks (Blocks 3, 4, and 5). In the Uncorrelated Naming Condition, however, there were significant increases between nearly all Blocks as the experiments progressed, excepting between Blocks 4 and 5 (Block 1 to 2, $p < 0.05$; Block 1 and Blocks 3, 4, and 5, $p < 0.001$; Block 2 and Blocks 3, 4, and 5, $p < 0.001$; Block 3 and Block 4, $p < 0.01$, and Block 3 and Block 5, $p < 0.001$).

When examining reaction time scores, while analyses of variance showed only a marginal effect of Naming Condition (Mean for Correlated = 1449.7 ms; Uncorrelated Mean = 1381 ms; $F(1, 74) = 2.18$, $p = 0.14$), there was a significant main effect of Block [$F(4, 296) = 343.25$, $p < 0.001$]. Post-hoc Tukey contrasts showed significant RT differences between Block 1 and all following blocks (Blocks 2, 3, 4, and 5, $p < 0.001$), and Block 2 and Blocks 4 and 5 ($p < 0.001$). Other comparisons were not significantly different.

There was also, interestingly, a significant interaction between Naming Condition and Block [$F(4, 252) = 4.00$, $p < 0.01$]. To unpack this interaction, reaction time scores were first separated by Naming Condition for follow-up ANOVAs. Significant effects were found for each Naming Condition [Correlated Naming Condition: $F(4, 160) = 143$, $p < 0.001$; Uncorrelated Naming Condition: $F(4, 136) = 209.2$, $p < 0.001$]. Additionally, the same relationships between Blocks held during post-hoc Tukey tests, when separated by Naming Condition: namely, there was a significant decrease in RT between the first and all following Blocks ($p < 0.001$ for all); all other contrasts were not significant.

Accuracy and RT scores were then divided by Trial Type (that is, type of phonological and/or visual overlap per trial) and subjected to ANOVAs. Due to the nature of the experimental design, analyses were separated by Naming Condition.

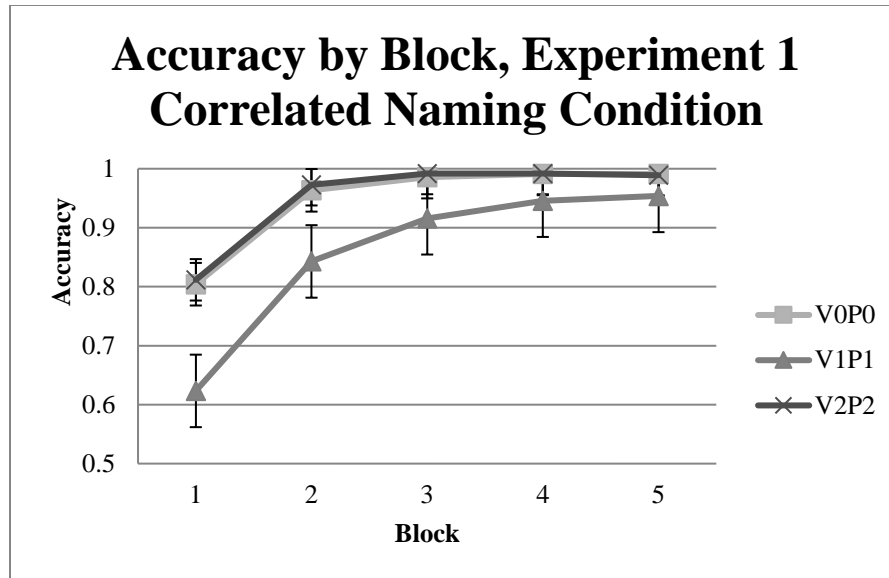


Figure 8. Accuracy, Correlated Naming Condition, Experiment 1. Note: V= visual overlap; P= phonological overlap. 1 indicates overlap is in the cap/first syllable; 2 indicates overlap in stem/second syllable. 0 indicates no overlap.

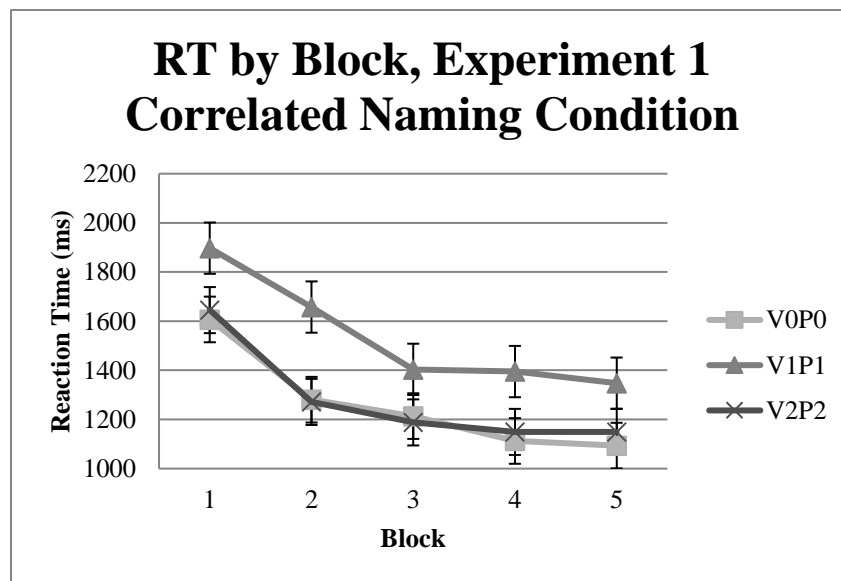


Figure 9. RT, Correlated Naming Condition, Experiment 1. Note: V= visual overlap; P= phonological overlap. 1 indicates overlap is in the cap/first syllable; 2 indicates overlap in stem/second syllable. 0 indicates no overlap.

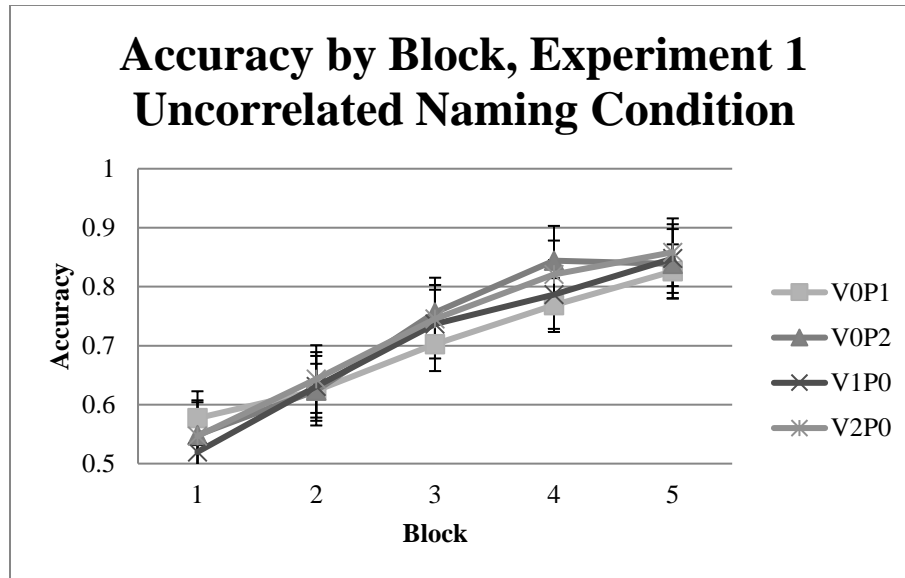


Figure 10. Accuracy, Uncorrelated Naming Condition, Experiment 1. Note: V= visual overlap; P= phonological overlap. 1 indicates overlap is in the cap/first syllable; 2 indicates overlap in stem/second syllable. 0 indicates no overlap.

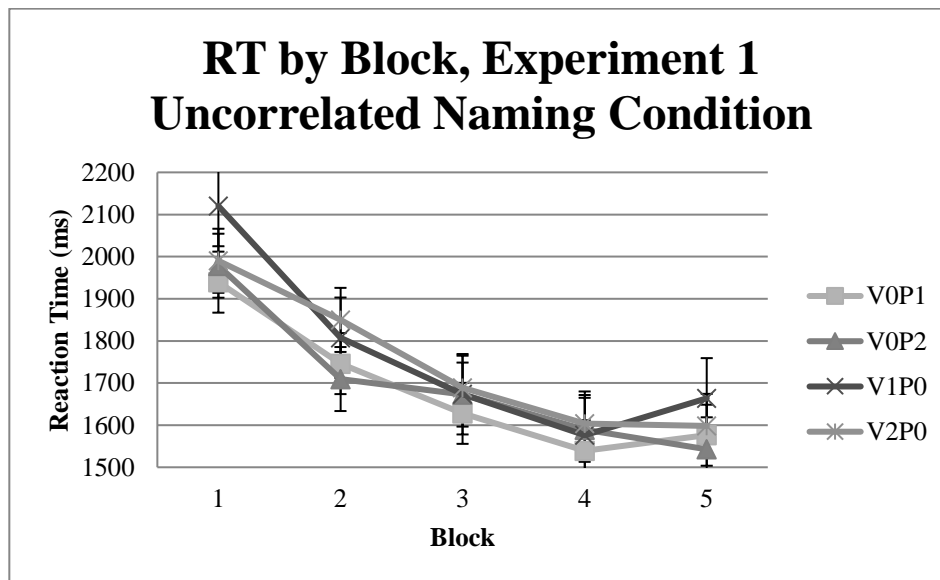


Figure 11. RT, Uncorrelated Naming Condition, Experiment 1. Note: V= visual overlap; P= phonological overlap. 1 indicates overlap is in the cap/first syllable; 2 indicates overlap in stem/second syllable. 0 indicates no overlap.

Figure 8 shows accuracy scores by Trial Type in the Correlated Naming Condition. There was a significant main effect of Trial Type on accuracy when collapsed across Blocks [$F(2, 572) = 38.62, p < 0.001$]. Post-hoc Tukey contrasts showed that this was driven by differences between cohort trials: they were both significantly less accurate than rhyme ($p < 0.001$) and no-overlap trials ($p < 0.001$).

In the Correlated Naming Condition, there was also a significant interaction between Trial Type and Block [$F(8, 560) = 8.22, p < 0.001$]. Follow-up ANOVAs of Block vs. accuracy, separated by Trial Type, showed that while there were still significant effects for every Trial Type [No overlap: $F(4, 160) = 81.88, p < 0.001$; Cohort: $F(4, 160) = 79.03, p < 0.001$; Rhyme: $F(4, 160) = 67.09, p < 0.001$], the relationships between Blocks differed by Trial Type--specifically, for cohort trials. For both no-overlap and rhyme trials, only the contrasts between Block 1 and following Blocks were significant (Block 1 and Blocks 2, 3, 4, and 5, $p < 0.001$). For cohort trials, not only was there significant increase in accuracy from the first Block throughout the experiment (Block 1 and Blocks 2, 3, 4, and 5, $p < 0.001$), this also held when comparing Block 2 against the following Blocks (Block 3, $p < 0.01$; Blocks 4 and 5, $p < 0.001$). Blocks 3, 4, and 5 were not significantly different from one another.

In the Uncorrelated Naming Condition (Figure 10), ANOVAs did not show a significant effect of Trial Type on task accuracy [$F(3, 586) = 0.89, p = 0.45$]; there was also no significant interaction between Trial Type and Block [$F(12, 570) = 1.09, p = 0.37$]. Additionally, for the Uncorrelated Naming Condition RT (Figure 11), there was neither a significant effect of Trial Type [$F(3, 586) = 1.67, p = 0.17$], nor a significant interaction between Trial Type and Block [$F(12, 570) = 0.43, p = 0.95$].

Discussion

As anticipated, the relationship between the names and features of an item influenced how quickly, and how well, participants learned the names of the items, even without instruction. Items whose names were indicative of their features were learned faster, and more correctly, than items whose names were not. Lexical competition effects, established in previous experiments (Magnuson et al., 2003), were replicated.

The next steps are to implement the paradigm used in this experiment in a task of phonological-orthographic learning, and compare individual differences results between the two experiments.

Chapter 3: Experiments 2a and 2b

In my first experiment, I examined whether performance scores on standardized tests of language ability or visual and language-related memory tasks could predict readers' ability to link new words to concrete visual objects. I focused on phonological-semantic learning, as a starting place for lexical learning. In my second experiment, I extended that line of questioning by examining phonological-orthographic learning, replacing the pictures of mushrooms from the first experiment with a novel orthography. Additionally, comparing experiment and individual differences assessment results from this experiment to the previous one could shed light on whether weaknesses in reading and reading comprehension stem from weaknesses in orthographic-phonological connections specifically, or in creating cross-modal connections, more generally.

Experiment 2a: Participants

Sixty-four University of Connecticut undergraduates were participants in the experiment. All participants were native, monolingual English speakers who reported normal hearing and normal or corrected-to-normal vision.

Experiment 2a: Apparatus and Materials

Assessments

The same five assessments from Experiment 1 were used to measure individuals' abilities in linguistic and para-linguistic domains: the Reading Span Task ((RS); verbal working memory; van den Noort et al., 2008), tests of sight-word and pseudoword reading efficiency (TOWRE SWE and PDE, respectively; Torgeson, Wagner & Rashotte, 1997), and tests of face, object, and Spoken Word memory of our own construction (described in Chapter 2).

Artificial lexicon experiment

Participants were to learn the names of nine novel orthographic character combinations. The orthographic characters, taken from Yoncheva, Blau, Maurer, and McCandliss, 2010, were formally equivalent to the mushroom stimuli used in Experiment 1. Instead of being constructed from three caps and three stems, these characters were formed from three left- and three right components. These components overlapped spatially, forming nine continuous, but parseable items (see Figure 12 and Figure 13 for illustrations of items and item-name pairings): that is, while the orthographic stimuli are constructed of solid lines, it is apparent that they are composed of recurring left- and right-side components. However, the left and right letters were partially superimposed to mask their componential nature, and mimic the continuous items from
















	 (&#	 '%#	 \$%&#
 !'&#	 !'(&#	 !' "%#	 !' \$%&#
 +, #	 +, (&#	 +, ' %#	 +, \$%&#
) * &#	) * (&#	) * ' %#	) * \$%&#

Figure 12. Sample sound-item pairings for stimuli in the Correlated Condition in Experiment 2a. Note that actual first- and second-syllable labels were counterbalanced across participants, so that one participant's “/dova/” may have been another's “/gule/,” but all relationships between left and right characters remained constant.

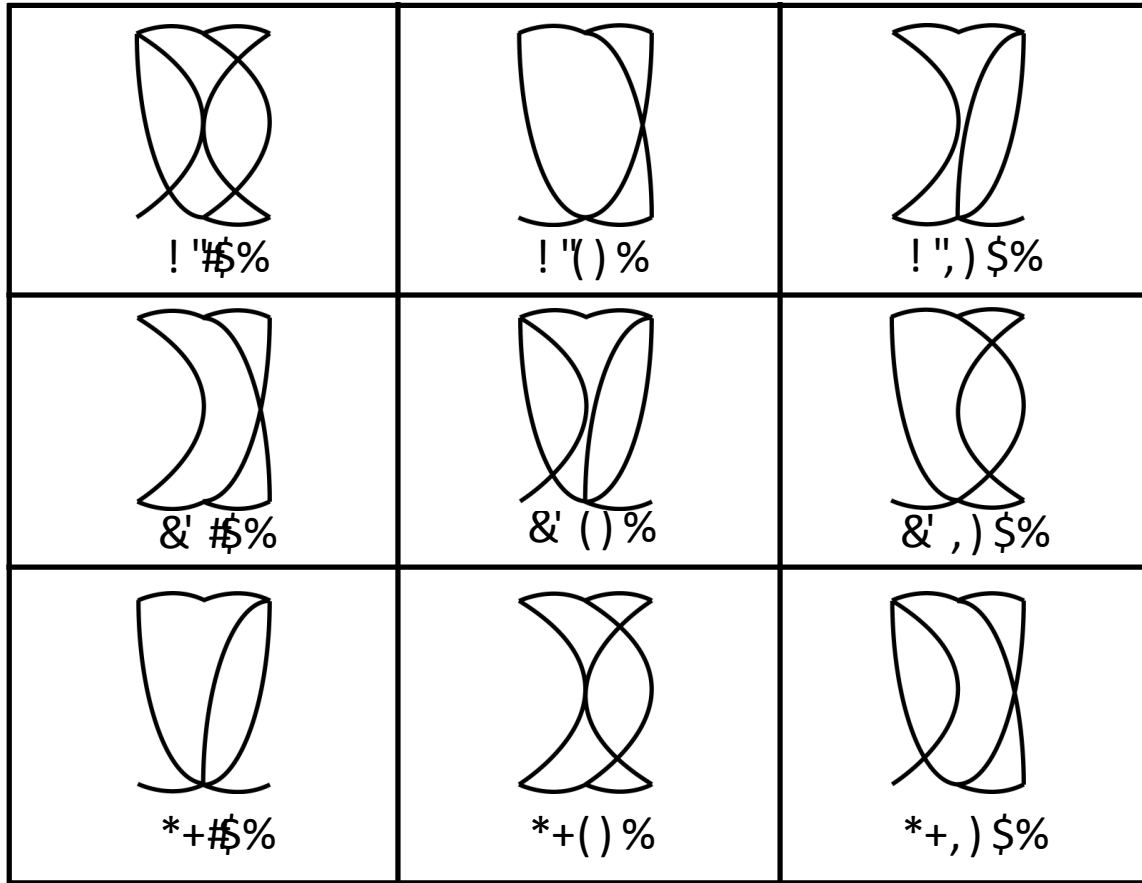


Figure 13. Sample sound-item pairings for stimuli in the Uncorrelated Condition in Experiment 2a. Note that actual first- and second-syllable labels were counterbalanced across participants, so that one participant's “/dole/” may have been another's “/piva/,” but all relationships between the mushroom caps and stems remained constant.

Experiment 1. The same phonological stimuli as in Experiment 1 were used, with the same naming schemes (left-right instead of cap-stem), creating Correlated ($n=29$) and Uncorrelated ($n=35$) Naming Conditions.

Experiment 2a: Procedure

The procedure was the same as that of Experiment 1: the testing session began with the assessments and the exposure phases of the old/new tasks. This was followed by the artificial lexicon experiment and then the test phases of the old/new tasks. As in Experiment 1, participants were assigned randomly to the experimental Naming Conditions, and were naive to possible item-name relationships in either Naming Condition. They were simply told they would learn the names of the objects, in a 2-alternative forced choice task.

Experiment 2a: Results

Accuracy and Reaction Time

Both Conditions showed an increase in accuracy (and corresponding decrease in RT) across the experiment and in each Naming Condition (see Figure 14 and *Figure 15*). Cohort effects were clear in each Naming Condition, evidenced by lower accuracy scores (in both Naming Conditions, Figure 16 and Figure 18) and slower response times (in the Correlated Naming Condition, *Figure 17*, though not for the Uncorrelated, Figure 19) for cohort trials when compared to the other trial types.

Analyses of variance. As in Experiment 1, to preserve a normal distribution of scores, accuracy scores were analyzed under a logit transformation, while reaction time scores were analyzed under a logarithmic transformation. Accuracy ANOVAs did not show a significant main effect of Naming Condition (Mean Correlated = 0.94; Mean Uncorrelated = 0.81; [$F(1, 62) = 2.95, p < 0.10$]). There was, however, a main effect of Block [$F(4, 248) = 241.29, p < 0.001$], though the

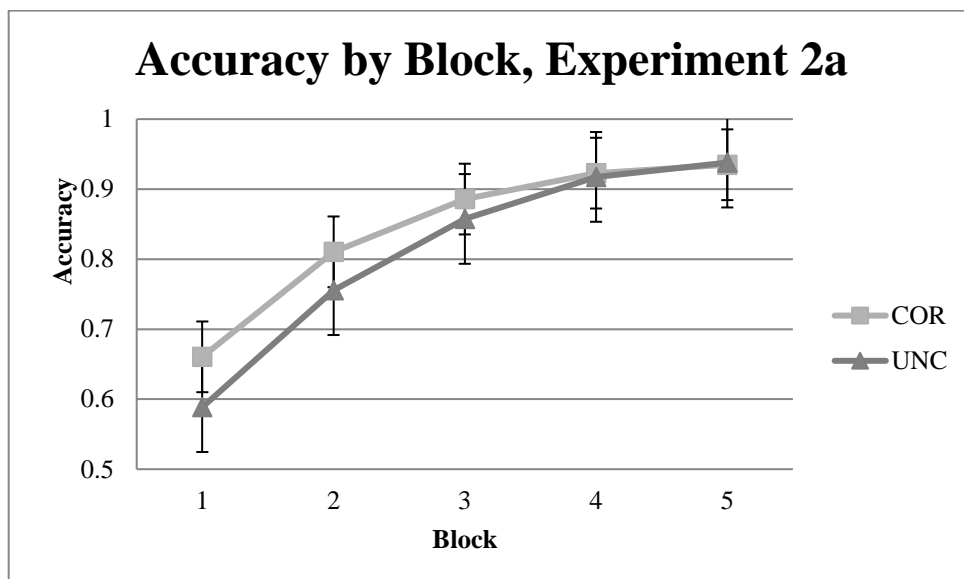


Figure 14. Accuracy scores for Experiment 2a. Note: COR= Correlated Naming Condition; UNC= Uncorrelated Naming Condition.

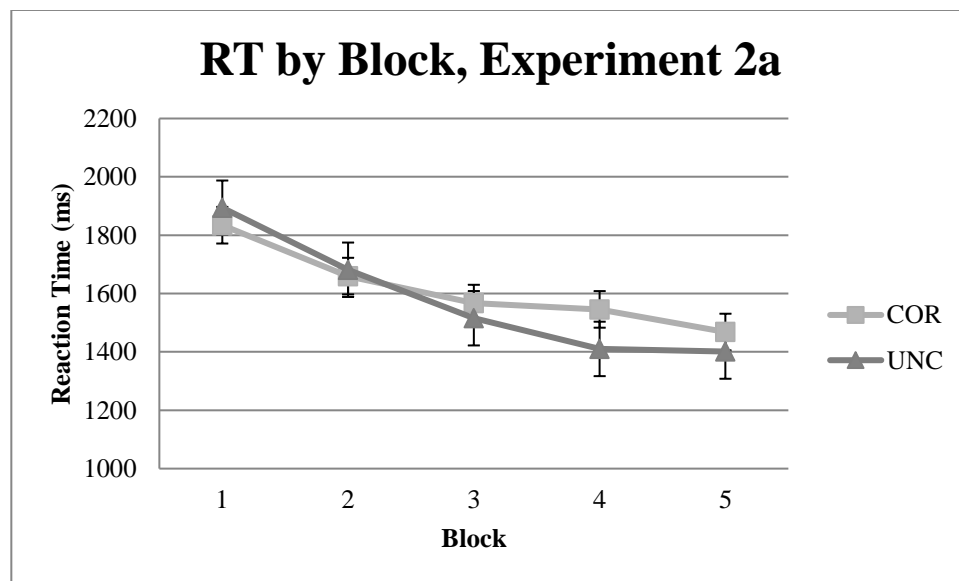


Figure 15. Reaction time for Experiment 2a. Note: COR= Correlated Naming Condition; UNC= Uncorrelated Naming Condition.

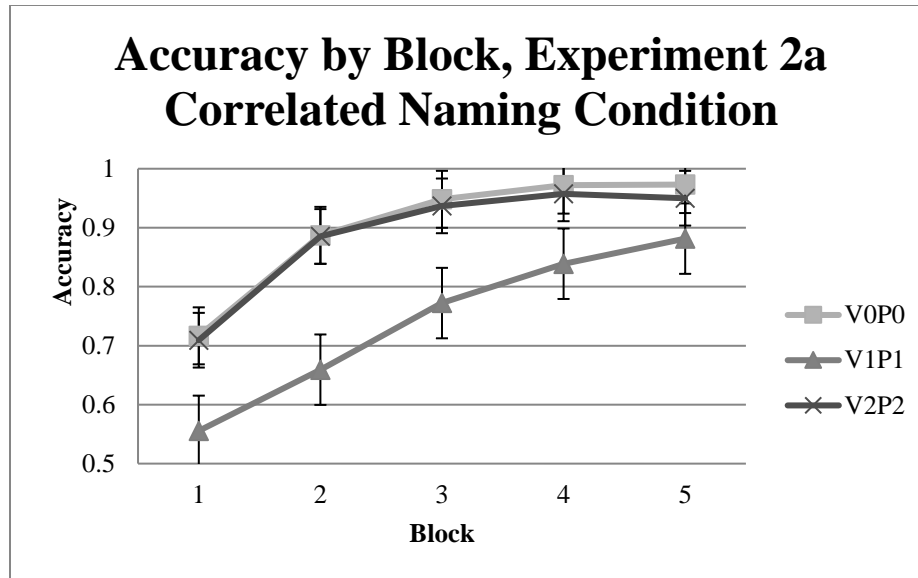


Figure 16. Accuracy, Correlated Naming Condition, Experiment 2a. Note: V= visual overlap; P= phonological overlap. 1 indicates overlap is in the cap/first syllable; 2 indicates overlap in stem/second syllable. 0 indicates no overlap.

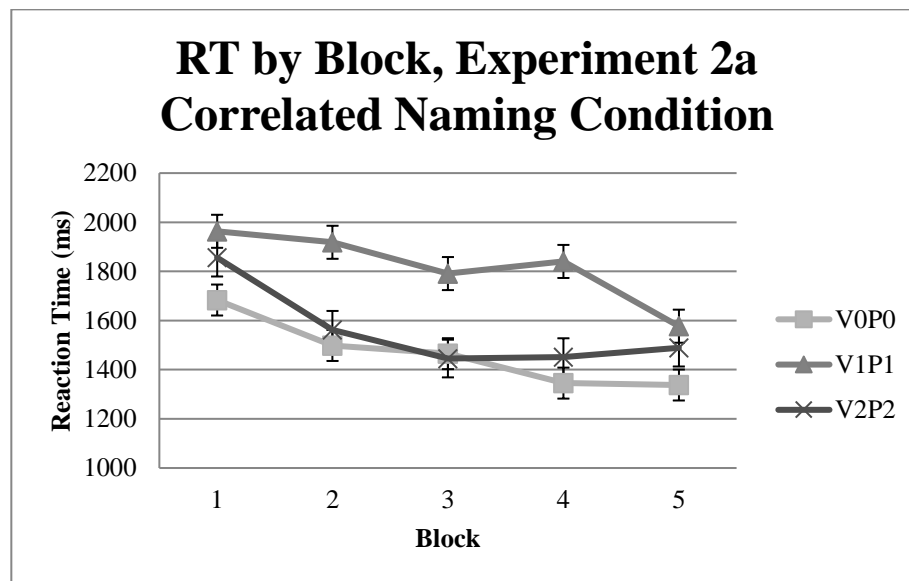


Figure 17. RT, Correlated Naming Condition, Experiment 2a. Note: V= visual overlap; P= phonological overlap. 1 indicates overlap is in the cap/first syllable; 2 indicates overlap in stem/second syllable. 0 indicates no overlap.

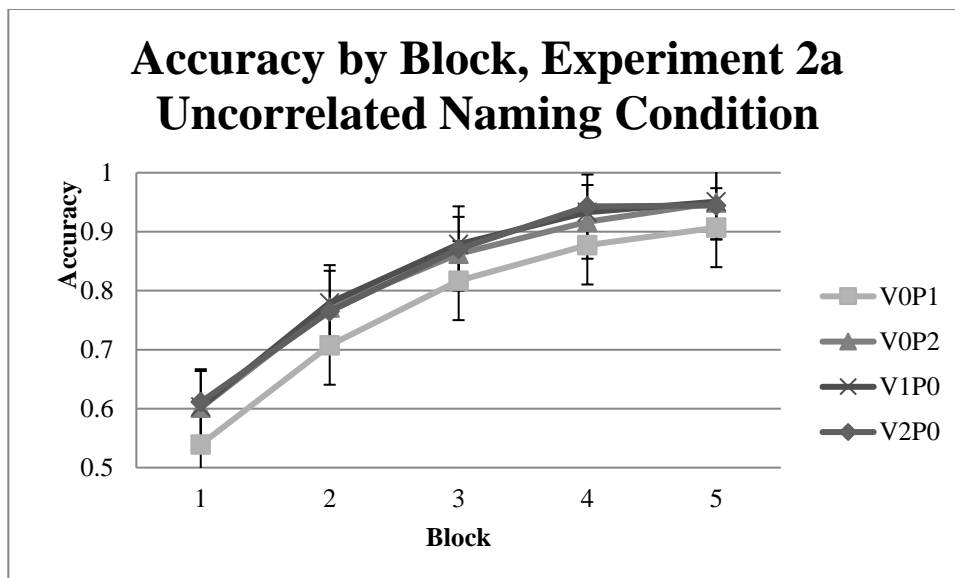


Figure 18. Accuracy, Uncorrelated Naming Condition, Experiment 2a. Note: V= visual overlap; P= phonological overlap. 1 indicates overlap is in the cap/first syllable; 2 indicates overlap in stem/second syllable. 0 indicates no overlap.

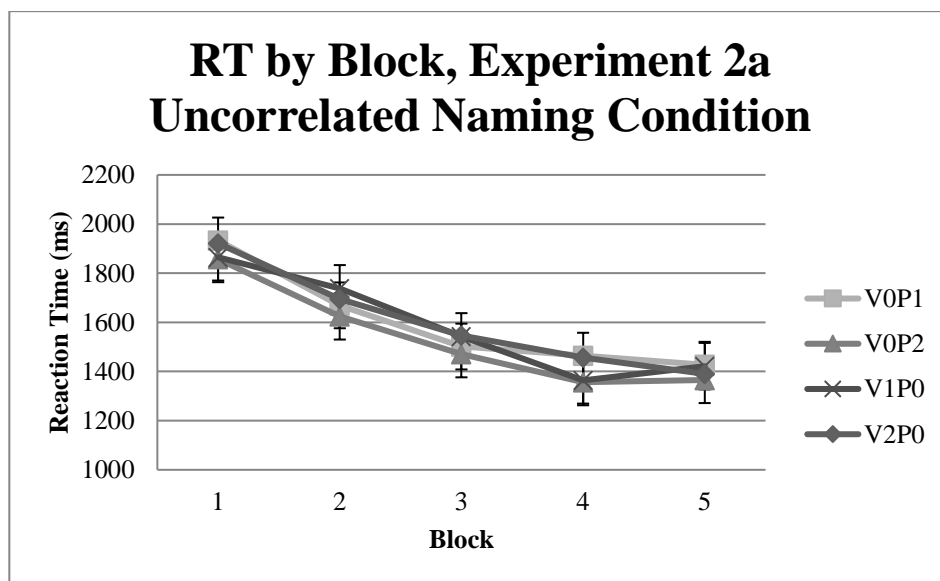


Figure 19. RT, Uncorrelated Naming Condition, Experiment 2a. Note: V= visual overlap; P= phonological overlap. 1 indicates overlap is in the cap/first syllable; 2 indicates overlap in stem/second syllable. 0 indicates no overlap.

interaction between Naming Condition and Block was not statistically significant [$F(4, 248) = 2.08, p < 0.10$]. Post hoc comparisons using Tukey HSD contrasts showed that all mean score comparisons for Block accuracy were significant ($p < 0.05$), except between Blocks 4 and 5. This shows that, despite the shallower accuracy slopes compared to the first experiment, participants were still making progress through nearly the entire experiment.

As with Accuracy scores, while ANOVAs did not show a significant main effect of Condition on reaction time performance (Mean Correlated = 1427 ms; Mean Uncorrelated = 1388.9 ms; [$F(1, 62) = 0.01, p = 0.91$]), there was a main effect of Block [$F(4, 248) = 409.15, p < 0.001$]. When post-hoc Tukey contrasts were performed by Block, significant RT differences were found between Block 1 and all following blocks (Blocks 2, 3, 4, and 5, $p < 0.001$). Other comparisons were not significantly different.

Additionally, as in the first Experiment, there was a significant interaction between Condition and Block [$F(4, 248) = 10.97, p < 0.001$]. Follow-up ANOVAs, separated by Naming Condition, showed that both Naming Conditions showed significant effects of RT by Block [Correlated Naming Condition: $F(4, 116) = 186.7, p < 0.001$; Uncorrelated Naming Condition: $F(4, 132) = 229.1, p < 0.001$], as well as similar patterns of between-Block differences. Post-hoc Tukey tests showed that, for each Naming Condition, only the contrasts between the first and following Blocks were significantly different (Block 1 versus Blocks 2, 3, 4, and 5, $p < 0.001$ for each contrast in each Naming Condition).

Next, accuracy and RT scores were examined by Trial Type (i.e., type of phonological and/or visual overlap per trial). Due to the nature of the experimental design, analyses were separated by Naming Condition.

Figure 16 shows accuracy scores by Trial Type in the Correlated Naming Condition. There was a significant main effect of Trial Type on accuracy when collapsed across Blocks [$F(2, 418) = 58.19, p < 0.001$]. Post-hoc Tukey contrasts showed that, as in the first experiment, this was driven by differences between cohort trials: they were both significantly less accurate than both rhyme ($p < 0.001$) and no-overlap trials ($p < 0.001$).

Additionally, there was a significant interaction between Trial Type and Block on task accuracy [$F(8, 406) = 3.47, p < 0.001$]. To unpack this interaction, follow-up accuracy ANOVAs, separated by Trial Type, were performed. As in the first experiment, while there were significant effects for every Trial Type [No overlap: $F(4, 116) = 91, p < 0.001$; Cohort: $F(4, 116) = 39.17, p < 0.001$; Rhyme: $F(4, 116) = 50.91, p < 0.001$], post-hoc Tukey contrasts showed that the relationships between Blocks differed for each Trial Type. Unlike in the first experiment, however, the pattern of contrasts between Blocks was unique for each Trial Type. For no-overlap trials, there was a significant increase in task accuracy from the first Block to all subsequent Blocks (Block 1 vs. Blocks 2, 3, 4, and 5, $p < 0.001$), as well as from the second Block to all following Blocks (Block 2 vs. Block 3, $p < 0.01$; vs. Blocks 4 and 5, $p < 0.001$). Cohort trials showed a longer period of significance across the experiment: in addition to the significant increases in accuracy from the first to final Blocks (Block 1 vs. Block 2, $p < 0.01$; all other contrasts, $p < 0.001$) and from the second Block through the final Block (Block 2 vs. Block 3, $p < 0.01$; vs. Blocks 4 and 5, $p < 0.001$), cohort trials also showed a significant increase in accuracy when comparing Block 3 to Block 5 ($p < 0.01$). Blocks 3 and 4, and 4 and 5, were not significantly different from one another. Finally, when examining rhyme trials, while the contrasts with Block 1 were consistent with those of other Trial Types (Block 1 vs. Blocks 2, 3, 4, and 5, $p < 0.001$), the second Block was only marginally different from Block 3 ($p = 0.9$),

though it was significantly less accurate than Blocks 4 ($p < 0.01$) and 5 ($p < 0.05$). No other contrasts were significant.

In the Uncorrelated Naming Condition (Figure 18), there was also a significant effect of Trial Type on accuracy when collapsed across Blocks [$F(3, 643) = 5.27, p < 0.01$]. Post-hoc Tukey contrasts showed that this was driven by trials where targets shared a first syllable; they were significantly less accurate than trials with second-syllable overlap ($p < 0.05$) as well as trials with either left-side or right-side overlap ($p < 0.01$ for each). There was not a significant interaction between Trial Type and Block on task accuracy [$F(12, 627) = 0.26, p = 0.99$].

Figure 17 shows reaction time scores by Trial Type in the Correlated Naming Condition. There was a significant main effect of Trial Type on task RT when collapsed across Blocks [$F(2, 406) = 33.16, p < 0.001$]; post-hoc Tukey tests showed that, as with accuracy, this was driven by differences between cohort trials: they were both significantly slower than both rhyme ($p < 0.001$) and no-overlap trials ($p < 0.001$).

Finally, when examining Uncorrelated Naming Condition RT performance (Figure 19), ANOVAs of RT vs. Trial Type and Block showed neither a main effect of Trial Type [$F(3, 627) = 2.03, p = 0.11$] nor a significant interaction [$F(12, 627) = 0.46, p = 0.94$].

Experiment 2a: Discussion

My second experiment extended the paradigms of the first, replacing phonological-visual/semantic pairings with phonology-orthography pairings—and, incidentally, increasing the stimulus difficulty. Compared to the first experiment, accuracy in the Correlated Naming Condition of the current experiment was lower, and accuracy in the Uncorrelated Naming

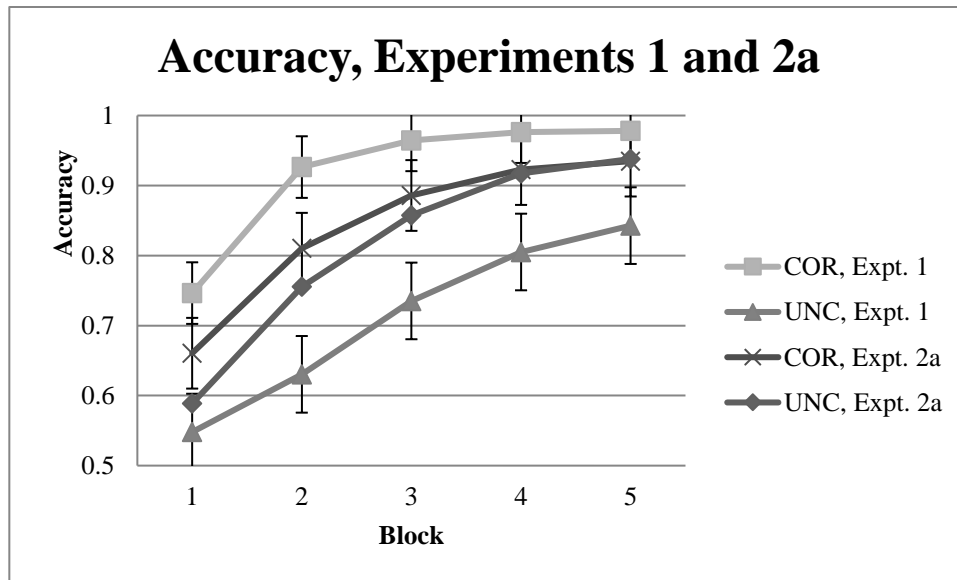


Figure 20. Accuracy scores for Experiments 1 and 2a. Note: COR= Correlated Naming Condition; UNC= Uncorrelated Naming Condition.

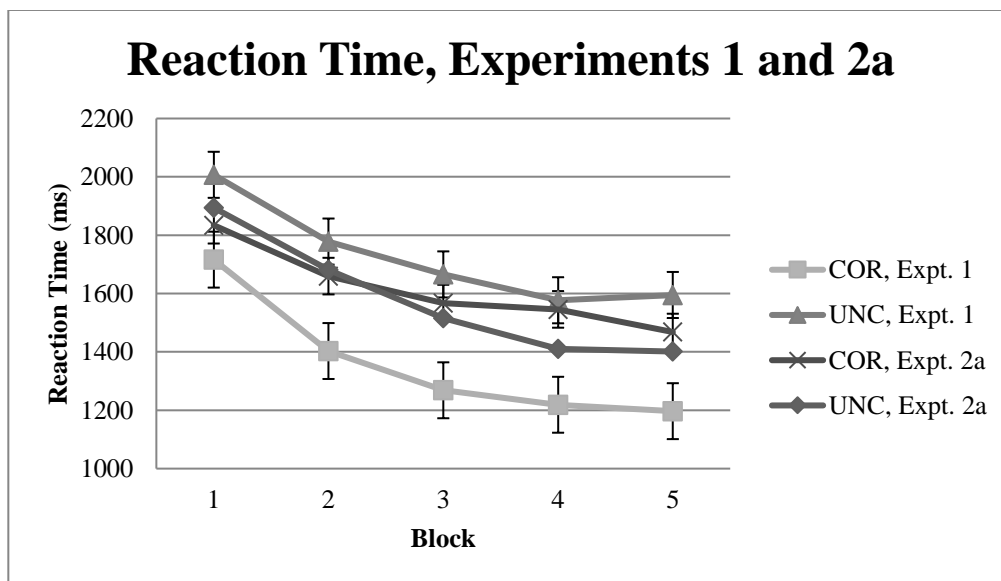


Figure 21. Reaction time for Experiments 1 and 2a. Note: COR= Correlated Naming Condition; UNC= Uncorrelated Naming Condition.

Condition was higher; this led to there being no significant difference in accuracy scores between the experimental Naming Conditions (see Figures 20 and 21).

This relative lack of overall difference between Naming Conditions for the current experiment may be due to different strategies used in response to the stimuli. Given that participants are told to learn the names of items, it is expected that they would try to detect consistent name-feature pairings. In the first experiment, the caps and stems of the mushrooms are easily parseable; pairing a syllable to a feature in the Correlated Naming Condition is not only easy, but also quickly rewarded. In the Uncorrelated Naming Condition of the first experiment, such a strategy would not be rewarded—though the items still look just as parseable, leading to a major cost in the first Blocks of the experiment from which participants did not completely recover.

In the second experiment, it seems that participants may be unable to employ componential learning strategies even in the Correlated Naming Condition, which would explain the diminished difference in accuracy and RT performance between Naming Conditions. While the novel orthographic stimuli are parseable (and are formally equivalent to the mushrooms in terms of component structure), it appears that regardless of Naming Condition, participants are simply learning the items as single characters, making performance in either Naming Condition the same: there was less advantage from the componential naming scheme in the Correlated Naming Condition, and less *disadvantage* from the perfectly uninformative naming scheme on the Uncorrelated Naming Condition. In other words, the difficulty encountered in breaking items down into their component parts not only prevents the benefits of an informative naming scheme, but it also prevents the costs of trying to break down the components when trying to learn the relationships in an uninformative naming scheme. This makes comparison of individual

differences correlations between the two experiments difficult, as participants appear to be using different strategies between experiments. With these results in mind, then, it follows that object recognition would be one of the most frequent predictors of performance in the orthographic task—participants' learning strategies in this task seems to be drawing on general visual skill.

Given that it is likely that participants simply did not pick up on the componential nature of the stimuli in this experiment, my next steps were to examine whether introducing "componential awareness"—in other words, alerting participants to the idea that the figures they would be learning were, in fact, composed of two contiguous characters—would influence the participants' learning strategies, and make comparisons between the phonological-visual/semantic and phonological-orthographic experiments more clear.

Experiment 2b: Participants

Twenty-two University of Connecticut undergraduates were participants in the experiment. All participants were native, monolingual English speakers who reported normal hearing and normal or corrected-to-normal vision.

Methods and Procedure

The methods, materials, and procedures replicated those of Experiment 2a, with one notable exception: before the learning task began, participants were shown an informational packet that depicted the orthographic character components, both separately and superimposed (sans the naming scheme, which was counterbalanced across participants), and told explicitly that individual characters could be combined to create a single, two-syllable word. This was done to ensure that participants were aware of the items' componential nature, and consequently results might better mirror the learning strategies used in Experiment 1 (where the components were more obvious, and perhaps difficult to learn to *not* use, in the Uncorrelated Naming Condition).

Experiment 2b: Results

Accuracy and Reaction Time

Figure 22 and Figure 23 show average accuracy and reaction times across the experiment. Both Naming Conditions increased in accuracy and decreased in RT across the experiment and in each Naming Condition. Cohort effects were apparent in each Naming Condition, evidenced by lower accuracy scores (in both Naming Conditions) and slower response times (in the Correlated Naming Condition, Figure 25) for cohort trials when compared to the other trial types. Rhyme effects were less clear, and much smaller—in the Correlated Naming Condition, accuracy for rhyme trials was lower than unrelated trials, but only in Blocks 2 and 3 (Figure 24); reaction time was also slower for rhyme trials compared to unrelated trials, but only in the first, third, and fifth Block (and these latter two effects were quite small). In the Uncorrelated Naming Condition, differences in accuracy between non-cohort trials were difficult to parse (Figure 26), and reliable reaction time differences for any trial type even more so (Figure 27).

Analyses of variance. As in Experiments 1 and 2a, to preserve a normal distribution of scores, accuracy scores were analyzed under a logit transformation, while reaction time scores were analyzed under a logarithmic transformation. Accuracy and RT scores were then subjected to analyses of variance. As for Experiment 2a, while accuracy ANOVAs did not show a significant main effect of Naming Condition (Mean Correlated = 0.85; Mean Uncorrelated = 0.77; [$F(1, 19) = 2.76, p < 0.10$]), there was a main effect of Block [$F(4, 76) = 69.45, p < 0.001$]. Post hoc comparisons using Tukey HSD contrasts showed that mean score comparisons for Block accuracy between the first Block and all subsequent Blocks (Blocks 2, 3, 4, and 5), as well as the second block and all subsequent Blocks (Blocks 3, 4, and 5) were significant ($p < 0.001$); Blocks, 3, 4, and 5 did not significantly differ.

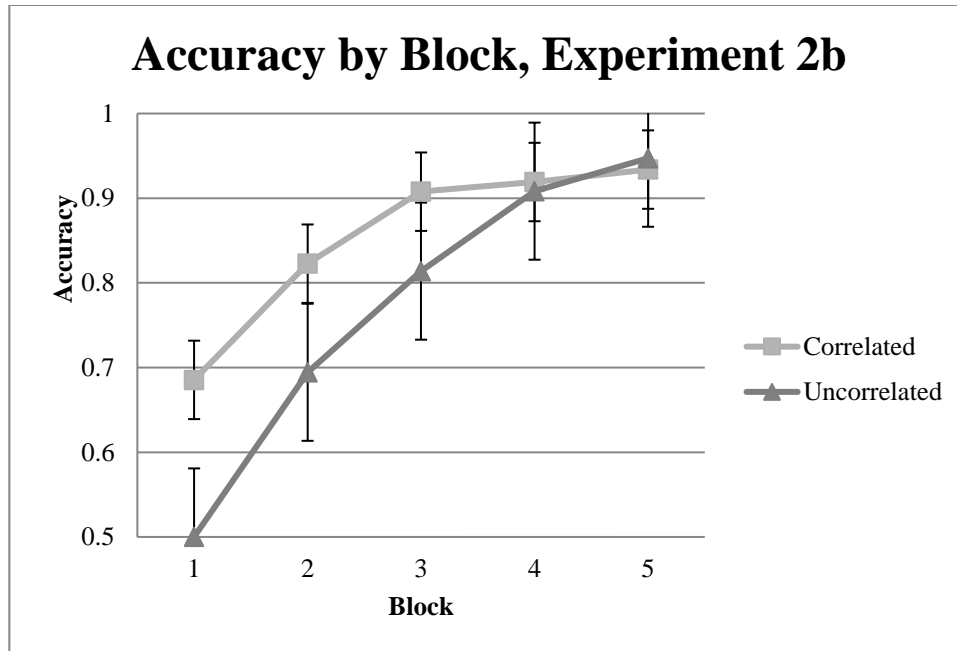


Figure 22. Accuracy scores for Experiment 2b. Note: COR= Correlated Naming Condition; UNC= Uncorrelated Naming Condition.

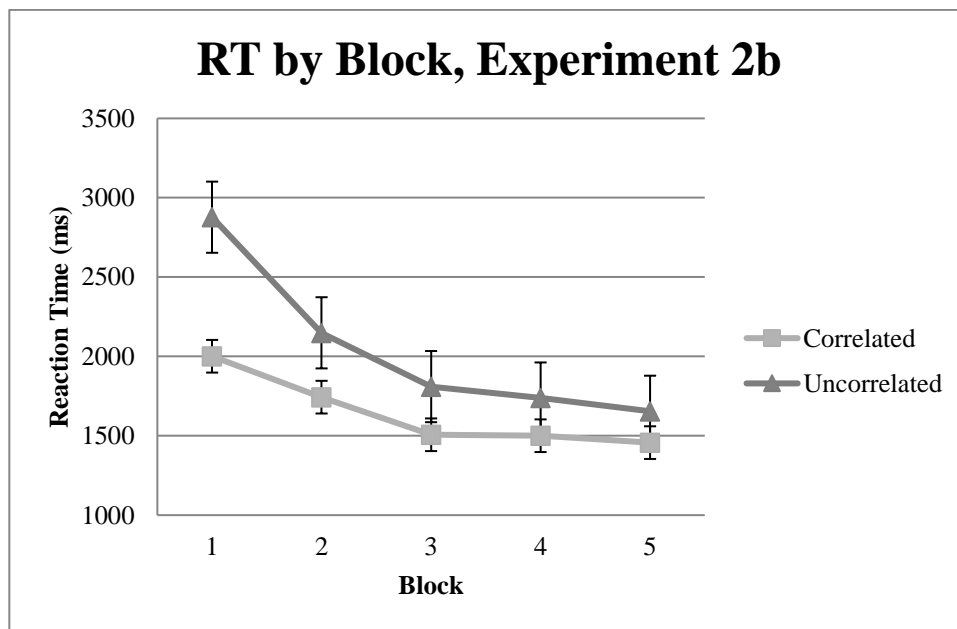


Figure 23. Reaction time for Experiment 2b. Note: COR= Correlated Naming Condition; UNC= Uncorrelated Naming Condition.

Additionally, there was a significant interaction between Naming Condition and Block [$F(4, 76) = 2.55, p < 0.05$]. Follow-up ANOVAs, separated by Naming Condition, showed that both Naming Conditions showed significant effects of accuracy by Block [Correlated Naming Condition: $F(4, 60) = 43.8, p < 0.001$; Uncorrelated Naming Condition: $F(4, 16) = 30.51, p < 0.001$]. However, the two conditions showed different patterns of accuracy increase by Block. Post-hoc Tukey tests in the Correlated Naming Condition showed significant contrasts for Block 1 and all subsequent Blocks (2, 3, 4, and 5, $p < 0.001$) as well as for Block 2 and all subsequent Blocks (Blocks 3, 4, and 5, $p < 0.001$). For the Uncorrelated Naming Condition, while Block 1 was still significantly less accurate compared to following Blocks (Block 2, $p < 0.05$; Block 3, 4, and 5, $p < 0.001$), Block 2 was only significantly less accurate than Blocks 4 and 5 ($p < 0.001$). Additionally, there was a significant contrast between Block 3 and 5 ($p < 0.001$).

As in with accuracy scores, an ANOVA on reaction time scores did not show a significant effect of Naming Condition (Mean Correlated = 1612.4 ms; Mean Uncorrelated = 1926 ms; [$F(1, 19) = 3.03, p < 0.10$]). There was a significant main effect of Block [$F(4, 76) = 12.240, p < 0.001$]; the interaction between Naming Condition was not statistically significant either [$F(4, 76) = 0.88, p < 0.10$]. To unpack the main effect of Block, RT scores were subjected to post-hoc Tukey HSD contrasts. The mean score of Block 1 was shown to be significantly slower than Blocks 3 ($p < 0.01$), 4 ($p < 0.001$), and 5 ($p < 0.001$); additionally, Block 2 was shown to be significantly slower than Blocks 4 and 5 ($p < 0.05$ and $p < 0.01$, respectively). No other contrasts were significant.

Accuracy and RT scores were divided by Trial Type (type of phonological and/or visual overlap per trial) and subjected to ANOVAs by Block; due to the nature of the experimental design, analyses were separated by Naming Condition.

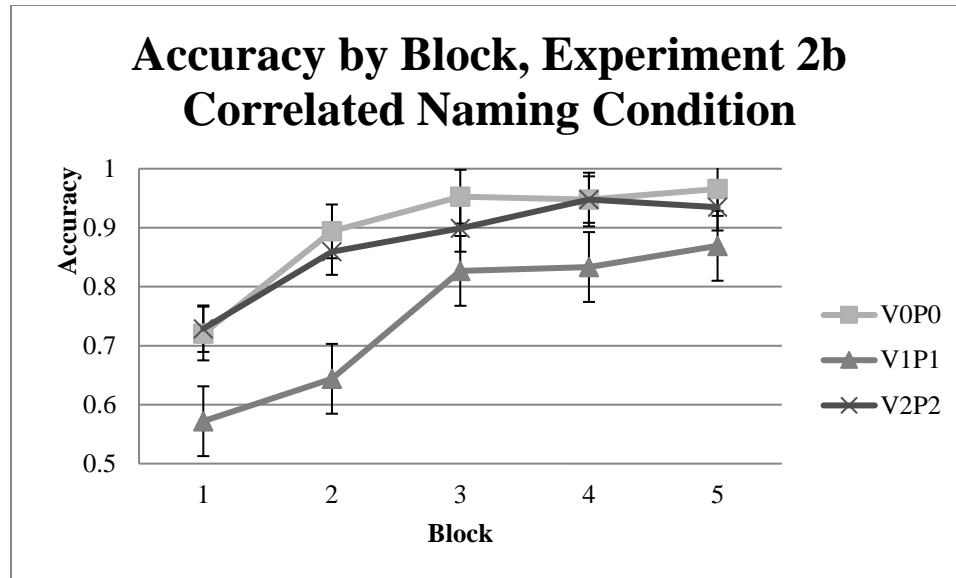


Figure 24. Accuracy, Correlated Naming Condition, Experiment 2b. Note: V= visual overlap; P= phonological overlap. 1 indicates overlap is in the cap/first syllable; 2 indicates overlap in stem/second syllable. 0 indicates no overlap.

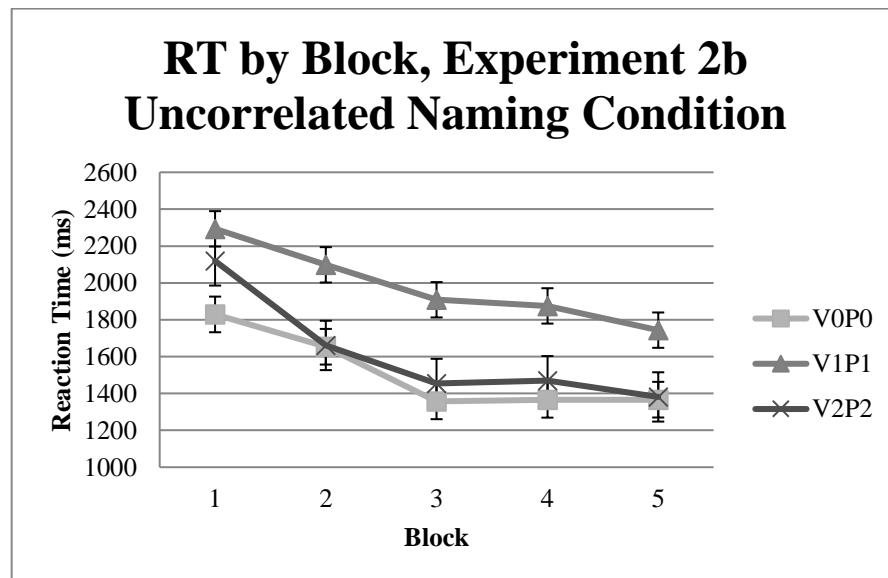


Figure 25. RT, Correlated Naming Condition, Experiment 2b. Note: V= visual overlap; P= phonological overlap. 1 indicates overlap is in the cap/first syllable; 2 indicates overlap in stem/second syllable. 0 indicates no overlap.

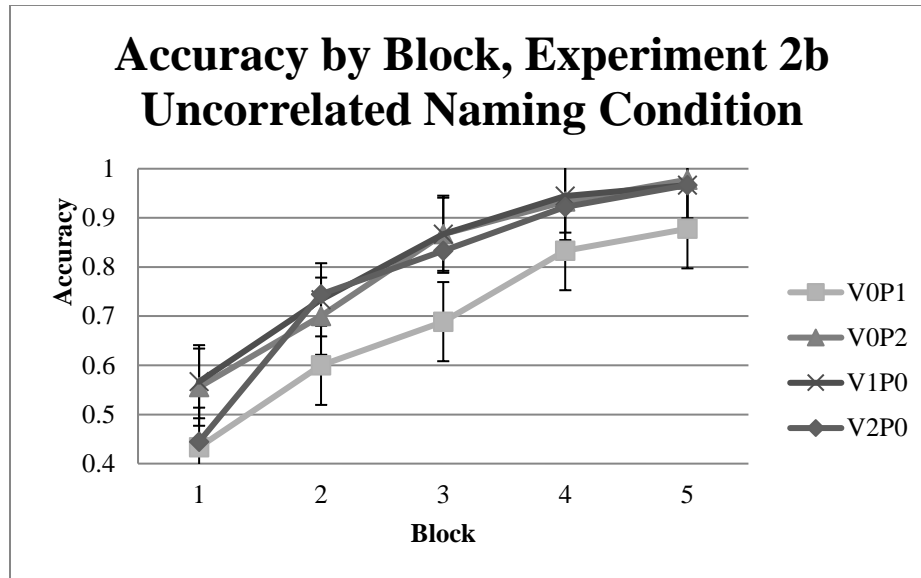


Figure 26. Accuracy, Uncorrelated Naming Condition, Experiment 2b. Note: V= visual overlap; P= phonological overlap. 1 indicates overlap is in the cap/first syllable; 2 indicates overlap in stem/second syllable. 0 indicates no overlap.

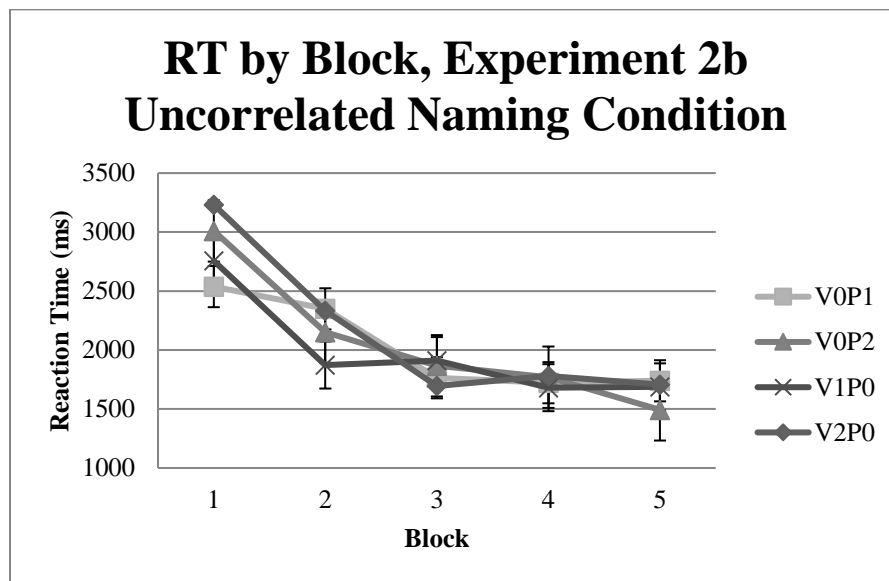


Figure 27. RT, Uncorrelated Naming Condition, Experiment 2b. Note: V= visual overlap; P= phonological overlap. 1 indicates overlap is in the cap/first syllable; 2 indicates overlap in stem/second syllable. 0 indicates no overlap.

Figure 24 shows accuracy scores by Trial Type in the Correlated Naming Condition. There was a significant main effect of Trial Type on accuracy when collapsed across Blocks [$F(2, 224) = 53.58, p < 0.001$]. As with previous experiments, post-hoc Tukey contrasts showed that this was driven by differences between cohort trials, which were significantly less accurate than both rhyme ($p < 0.001$) and no-overlap trials ($p < 0.001$).

There was also a significant interaction between Trial Type and Block on Correlated Naming Condition Accuracy [$F(2, 224) = 2.35, p < 0.05$]. Follow-up accuracy ANOVAs, separated by Trial Type, showed that while the effect of Block remained significant for each Trial Type [No-overlap: $F(4, 64) = 30.42, p < 0.001$; Cohort: $F(4, 64) = 36.14, p < 0.001$; Rhyme: $F(4, 64) = 16.48, p < 0.001$], the relationships between accuracy scores and Blocks was unique for each Trial Type. For both no-overlap and rhyme trials, post-hoc Tukey contrasts showed that there was a significant increase in accuracy scores for the first Block and all following Blocks (2, 3, 4, and 5, $p < 0.001$ for all cases). Each Trial Type only had one other significant contrast: for no-overlap trials, it was between Blocks 2 and 5 ($p < 0.05$), and for rhyme trials, it was between Blocks 2 and 4 ($p < 0.05$); all other contrasts were insignificant. For cohort trials, the accuracy-by-Block relationships differed. While there was no significant difference between Blocks 1 and 2 ($p = 0.14$), there was a significant difference between Blocks 1 and Blocks 3, 4, and 5, as well as between Blocks 2, 3, 4, and 5 ($p < 0.001$ in all cases). No other contrasts were significant.

In the Uncorrelated Naming Condition (Figure 26), there was also a significant effect of Trial Type on accuracy when collapsed across Blocks [$F(3, 76) = 7.69, p < 0.001$]. Post-hoc Tukey contrasts revealed that while the main effect was significant, only one contrast showed even marginal significance (left-side overlap and first-syllable overlap, $p = 0.09$). The interaction between Trial Type and Block was not significant [$F(3, 76) = 0.42, p = 0.95$].

There was a significant effect of Trial Type on RT when collapsed across Blocks in the Correlated Naming Condition (Figure 25, [$F(2, 224) = 22.029, p < 0.001$]). Post-hoc Tukey HSD contrasts showed that cohort trials were significantly slower than trials with no overlap as well as rhyme trials ($p < 0.001$). The interaction between Trial Type and Block was not significant [$F(8, 224) = 0.36, p = 0.94$].

In the Uncorrelated Naming Condition (Figure 27), there was neither a significant main effect of Trial Type on RT [$F(3, 76) = 0.45, p = 0.72$] nor a significant interaction between Trial Type and Block [$F(12, 76) = 0.60, p = 0.83$].

While this addition of componential awareness seems to have aided replication of established lexical findings consistent with use of a visual world paradigm, overall accuracy and reaction time relationships between the Naming Conditions have replicated the results of Experiment 2a. Unfortunately, this suggests that participants may still be favoring holistic strategies over componential ones, and more action must be taken to be able to confidently compare between the phonological-visual and phonological-orthographic experiments.

Experiment 2b: Discussion

Averaged accuracy and reaction time results did not improve upon those of Experiment 2a, which means that further manipulation of componential awareness is needed if comparisons between the phonological-visual/semantic and phonological-orthographic versions of this experiment are to be made. Making the visual stimuli more obviously componential—as in, separating them—may prove helpful.

Chapter 4: Experiment 3

My third experiment seeks to improve upon previous paradigms designed to find out whether performance scores on standardized tests of language ability or visual and language-related memory tasks could predict readers' ability to link new words to concrete visual objects. As in Experiment 2, the following experiment focuses on phonological-orthographic learning. However, unlike in Experiment 2, visual stimuli used in this study are separated into two distinct figures, in hopes of avoiding potential segmentation difficulties that may have made it more difficult to confidently compare the outcomes from my phonological-orthographic experiments to results from my phonological-visual experiment.

Methods

Participants

Sixty-five University of Connecticut undergraduates were participants in the experiment. All participants were native, monolingual English speakers who reported normal hearing and normal or corrected-to-normal vision.

Apparatus and Materials

Assessments. The same five assessments from Experiments 1 and 2 were used to measure individuals' abilities in linguistic and nonlinguistic domains: the Reading Span Task (RS; verbal working memory; van den Noort et al., 2008), tests of single-word and pseudoword reading efficiency (TOWRE SWE and PDE, respectively; Torgeson, Wagner & Rashotte, 1997), and tests of Face, Object, and Spoken Word recognition of our own construction.

Artificial lexicon experiment. The experimental design was informed by Experiments 2a and 2b. Participants were to learn the names of nine novel orthographic character combinations. The orthographic characters, taken from Yoncheva, Blau, Maurer, & McCandliss, 2010, were constructed similarly to the mushroom stimuli in Experiments 2a and 2b: three unique left-side items were combined with three unique right-side items, creating nine item pairs. However, unlike in Experiments 2a and 2b, the left and right items were kept *separate*, so that the componential nature of the item pairs was apparent (see Figure 28 and Figure 29). The same phonological stimuli as in the Experiments were used, with the same naming schemes (left-right), creating Correlated (n= 32) and Uncorrelated (n= 33) Naming Conditions.
















	 le	 va	 sae
 pi	 pile	 piva	 pisae
 do	 dole	 dova	 dosae
 gu	 gule	 guva	 gusae

Figure 28. Sample sound-item pairings for stimuli in the Correlated Condition in Experiment 3. Note that actual first- and second-syllable labels were counterbalanced across participants, so that one participant's "/dova/" may have been another's "/gule/," but all relationships between left and right characters remained constant.










 pile	 piva	 pisae
 dole	 dova	 dosae
 gule	 guva	 gusae

Figure 29. Sample sound-item pairings for stimuli in the Uncorrelated Condition in Experiment 3. Note that actual first- and second-syllable labels were counterbalanced across participants, so that one participant's “/dole/” may have been another's “/piva/,” but all relationships between the mushroom caps and stems remained constant.

Procedure

The procedure was identical to that of Experiment 2b. The testing session began with the RS and TOWRE assessments and the exposure phases of the old/new tasks. This was followed by the artificial lexicon experiment and then the test phases of the old/new tasks.

As in Experiment 2b, we introduced "componential awareness" to the participants via an explanatory packet before the testing session. Participants were shown all of the possible visual pairings of the orthographic stimuli beforehand, regardless of assigned Naming Condition. Participants were assigned randomly to the experimental Naming Conditions, and were naive to possible naming relationships in the materials in either Naming Condition; they were simply told they would learn the names of the objects, in a 2-alternative forced choice task.

Results

Reaction Time and Accuracy

Figure 30 and Figure 31 show averaged accuracy and reaction time scores across the experiment. As with the previous experiments, for both Naming Conditions, accuracy increased, and reaction time decreased, across the experiment. The Correlated Naming Condition was much easier than the Uncorrelated Naming Condition for both accuracy and reaction time (see Figure 32 and Figure 34, and Figure 33 and Figure 35, respectively). Fortunately, this replicates the relationship between the two Naming Conditions in Experiment 1, making it possible to directly compare results between the two experiments.

Examination of reaction time and accuracy data by trial type showed evidence of cohort and rhyme effects in the Correlated Naming Condition (the former was apparent in both reaction time and accuracy; the latter was apparent in reaction time, but only suggested in accuracy performance). In the Uncorrelated Naming Condition, reaction time data suggests an increased

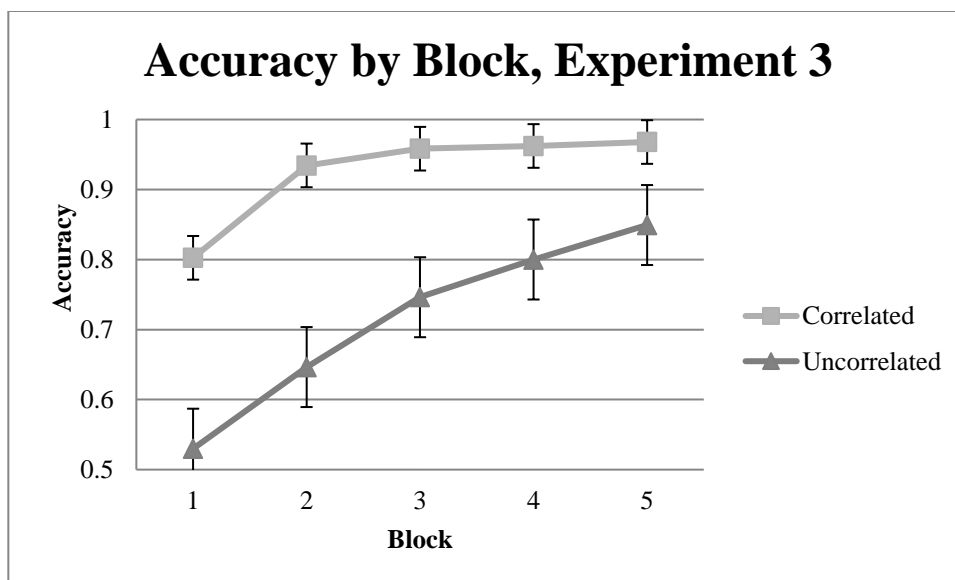


Figure 30. Accuracy scores for Experiment 3. Note: COR= Correlated Naming Condition; UNC= Uncorrelated Naming Condition.

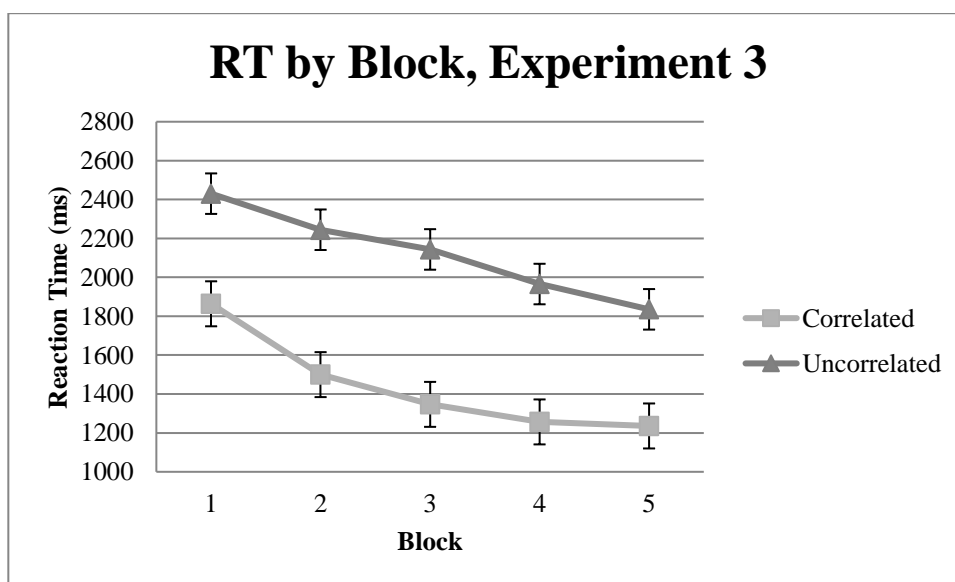


Figure 31. Reaction time for Experiment 3. Note: COR= Correlated Naming Condition; UNC= Uncorrelated Naming Condition.

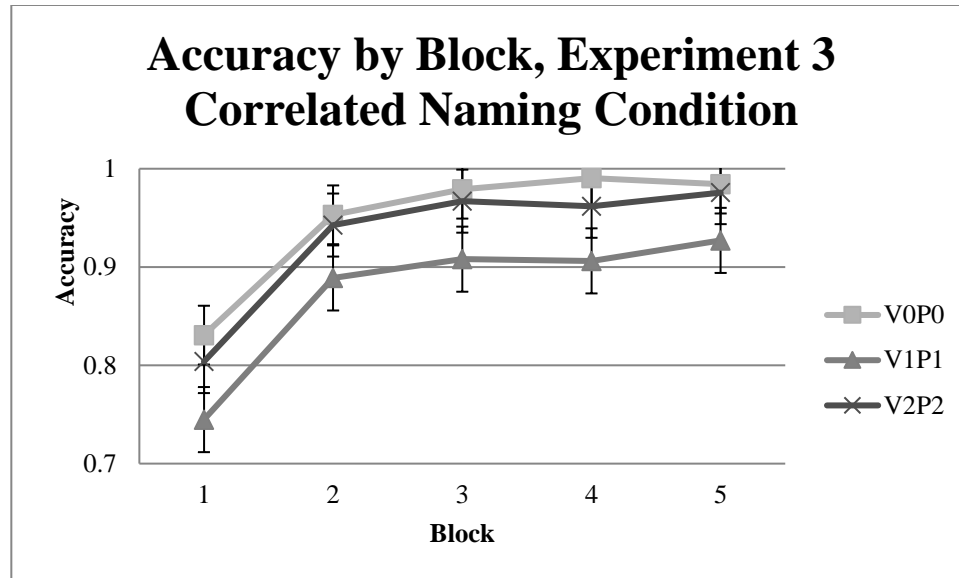


Figure 32. Accuracy, Correlated Naming Condition, Experiment 3. Note: V= visual overlap; P= phonological overlap. 1 indicates overlap is in the cap/first syllable; 2 indicates overlap in stem/second syllable. 0 indicates no overlap.

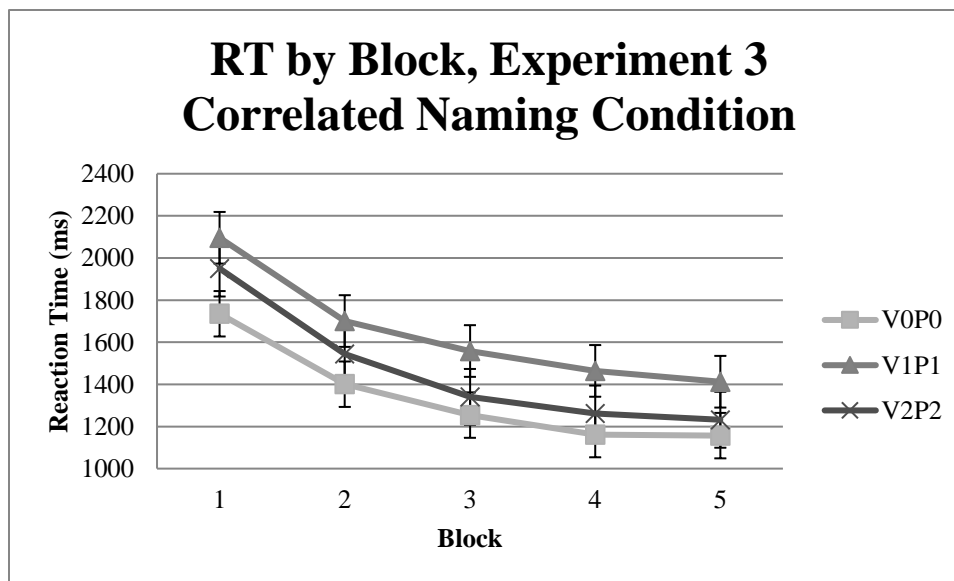


Figure 33. RT, Correlated Naming Condition, Experiment 3. Note: V= visual overlap; P= phonological overlap. 1 indicates overlap is in the cap/first syllable; 2 indicates overlap in stem/second syllable. 0 indicates no overlap.

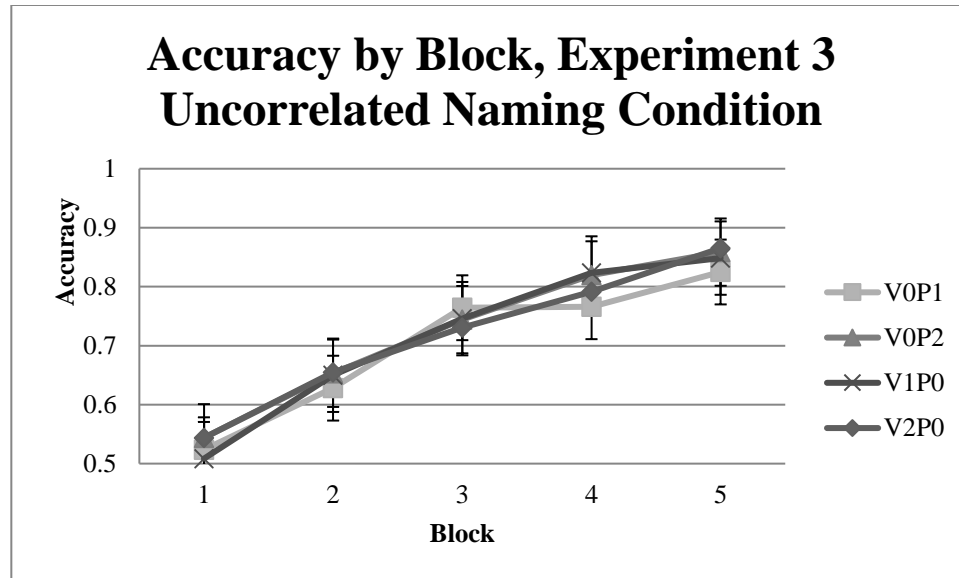


Figure 34. Accuracy, Uncorrelated Naming Condition, Experiment 3. Note: V= visual overlap; P= phonological overlap. 1 indicates overlap is in the cap/first syllable; 2 indicates overlap in stem/second syllable. 0 indicates no overlap.

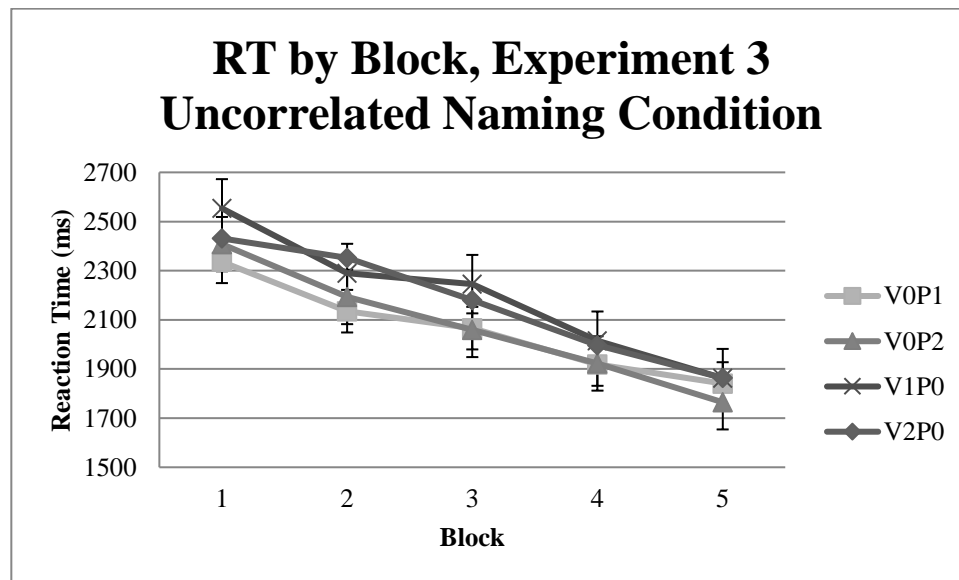


Figure 35. RT, Uncorrelated Naming Condition, Experiment 3. Note: V= visual overlap; P= phonological overlap. 1 indicates overlap is in the cap/first syllable; 2 indicates overlap in stem/second syllable. 0 indicates no overlap.

difficulty for visual overlap trials compared to phonological overlap trials, but the difference is minimal.

Analysis of variance. As in previous Experiments, to preserve a normal distribution of scores, accuracy scores were analyzed under a logit transformation, while reaction time scores were analyzed under a logarithmic transformation. Analyses of variance showed a main effect of Naming Condition on accuracy scores (Mean for Correlated = 0.93; Uncorrelated Mean = 0.71; $F(1, 63) = 95.26, p < 0.001$), as well as a main effect of Block [$F(4, 256) = 110.0, p < 0.001$], both corroborating the visual inspection of accuracy data, and replicating the accuracy results between Naming Conditions from Experiment 1. This shows that, compared to the previous stimulus-sound pairings, the separated orthography was the most salient. Post hoc comparisons (using Tukey HSD contrasts) showed increases in accuracy performance from Block 1 versus all proceeding Blocks (Blocks 2, 3, 4, and 5; $p < 0.001$) as well as a significant increase in Block 5 accuracy vs. Block 2 accuracy ($p < 0.01$). This also can be seen in Figure 30, where the slope for each Naming Condition shallows as the experiment progresses.

Additionally, there was also a significant interaction between Naming Condition and Block [$F(4, 252) = 8.17, p < 0.001$]. Follow-up ANOVAs by Naming Condition showed significant effects of accuracy by Block [Correlated Naming Condition: $F(4, 124) = 56.4, p < 0.001$; Uncorrelated Naming Condition: $F(4, 128) = 73.98, p < 0.001$]. Post-hoc Tukey HSD analyses showed two different patterns as the experiment progressed. In the Correlated Naming Condition, accuracy scores in the first Block were significantly lower than in subsequent Blocks (2, 3, 4, and 5, $p < 0.001$); additionally, accuracy in Block 2 was significantly lower than in Block 5 ($p < 0.01$); no other contrasts were significant. In the Uncorrelated Naming Condition, however, all contrasts were significant, meaning that accuracy scores significantly increased over the

course of the experiment [$p < 0.001$ for all contrasts, excepting between Blocks 3 and 4 ($p < 0.05$) and Blocks 4 and 5 ($p < 0.01$)].

RT ANOVAs showed main effects of both Naming Condition (Mean for Correlated = 1425.4 ms; Uncorrelated Mean = 2086 ms; [$F(1, 63) = 37.91, p < 0.001$]) and Block [$F(4, 256) = 52.77, p < 0.001$]. Post-hoc Tukey HSD contrasts showed significant differences in RT between Blocks 1 and all following Blocks (2, 3, 4, and 5; $p < 0.001$), as well as between Block 2 and Blocks 4 and 5 ($p < 0.001$). No other contrasts were statistically significant.

Additionally, there was a significant interaction between Naming Condition and Block [$F(4, 252) = 4.00, p < 0.01$]. Follow-up ANOVAs by Naming Condition showed significant effects of RT by Block [Correlated Naming Condition: $F(4, 124) = 54.22, p < 0.001$; Uncorrelated Naming Condition: $F(4, 128) = 14.79, p < 0.001$]. Post-hoc Tukey HSD contrasts showed significant differences in the Correlated Naming Condition for Block 1 and following Blocks (2, 3, 4, and 5, $p < 0.001$), as well as for Block 2 and following Blocks (3, $p < 0.05$; 4 and 5, $p < 0.001$). No other contrasts were significant for Correlated Naming Condition RT. In the Uncorrelated Naming Condition, a different pattern emerged: post-hoc Tukey contrasts were significant for Block 1 vs. Blocks 3 ($p < 0.01$), 4, and 5 ($p < 0.001$), Blocks 2 and Blocks 4 ($p < 0.05$) and 5 ($p < 0.001$), and Blocks 3 and 5 ($p < 0.01$).

Finally, accuracy and RT scores were divided by Trial Type (type of phonological and/or visual overlap per trial; analyses were separated by Naming Condition, due to the nature of the experimental design).

Figure 32 illustrates accuracy scores by Trial Type in the Correlated Naming Condition. There was a significant effect of Trial Type on accuracy when collapsed across Blocks [$F(2, 429) = 28.95, p < 0.001$]. Post-hoc Tukey contrasts showed that, as with the previous

experiments, this was driven by cohort trials: they were significantly less accurate than both rhyme and no-overlap trials ($p < 0.001$). There was no significant interaction between Trial Type and Block [$F(8, 429) = 0.15, p = 0.99$].

In the Uncorrelated Naming Condition (Figure 34), there was neither a main effect of Trial Type on task accuracy [$F(3, 608) = 1.24, p = 0.30$], nor a significant interaction between Trial Type and Block [$F(12, 608) = 0.85, p = 0.60$].

Reaction Time scores by Trial Type for the Correlated Naming Condition are in Figure 33. There was a significant effect of Trial Type on RT when collapsed across Blocks [$F(2, 429) = 32.39, p < 0.001$]. Post-hoc Tukey contrasts showed significant differences between all three Trial Types: in addition to replicating previous relationships of slower RT for cohort than rhyme and no-overlap trials ($p < 0.001$), RT for rhyme trials was also significantly slower than RT for no-overlap trials ($p < 0.05$). There was no significant interaction between Trial Type and Block [$F(8, 429) = 0.30, p = 0.97$].

Figure 35 shows RT scores by Trial Type in the Uncorrelated Naming Condition. There was a significant effect of Trial Type on RT when collapsed across Blocks [$F(3, 608) = 4.168, p < 0.01$]. Post-hoc Tukey contrasts showed that this was driven by the contrast between left-side overlap trials and phonological-overlap trials; the former were significantly slower than the latter ($p < 0.05$). The interaction between Trial Type and Block was not significant [$F(12, 608) = 0.36, p = 0.98$].

Discussion

In this experiment, we improved upon the stimulus design of our novel orthography from the previous chapter; the introduction of separated orthographic images, along with more explicit

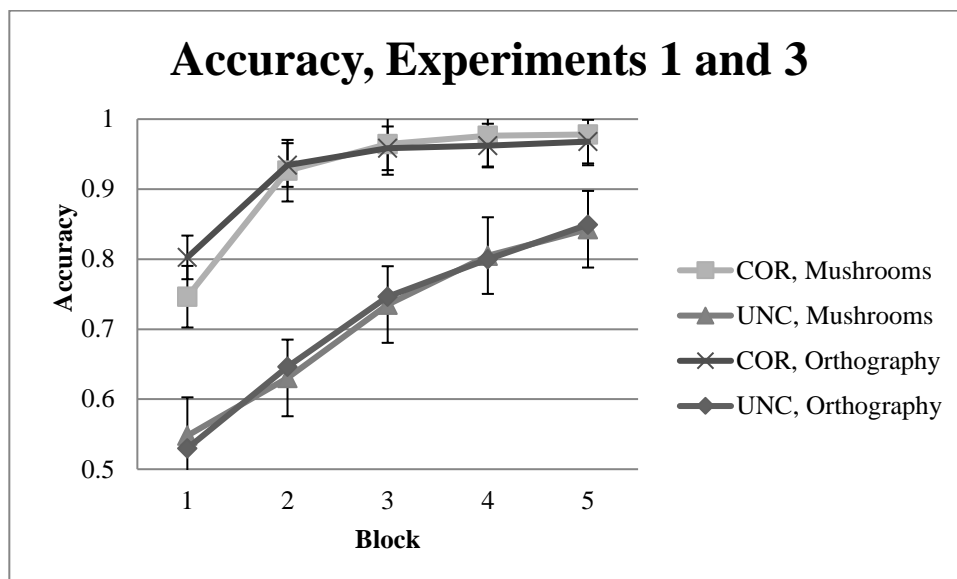


Figure 36. Accuracy scores for Experiments 1 and 3. Note: COR= Correlated Naming Condition; UNC= Uncorrelated Naming Condition.

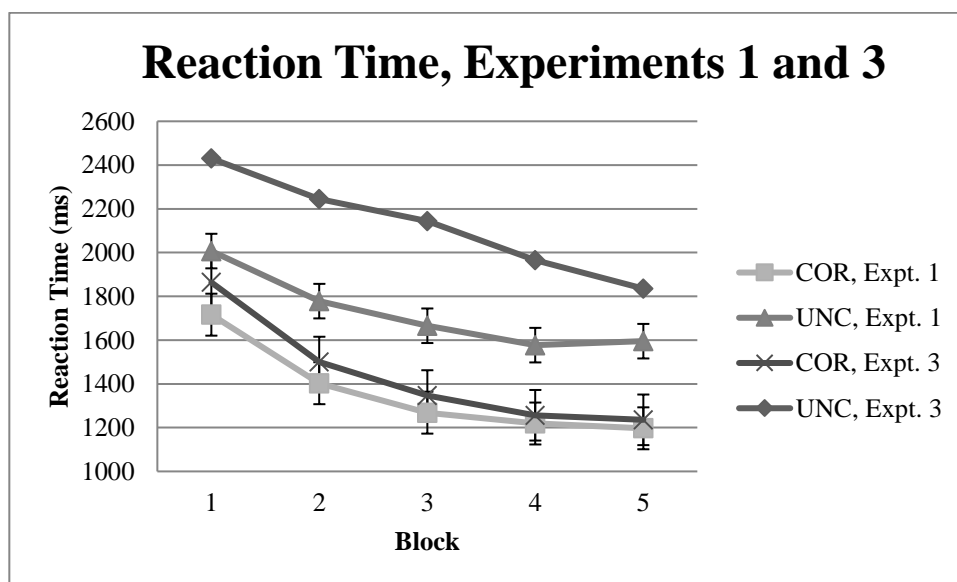


Figure 37. Reaction time for Experiments 1 and 3. Note: COR= Correlated Naming Condition; UNC= Uncorrelated Naming Condition.

instruction, led to results that more closely replicated results from the phonological-visual learning experiment. Figures 36 and 37 depict accuracy and RT results by Naming Condition for Experiments 1 and 3. Indeed, when evaluating the effect of Experiment, accuracy ANOVAs did not find a significant difference between the two experiments (Correlated Naming Accuracy [$F(1, 71) = 0.01, p < 1.0$]; Uncorrelated Naming Accuracy [$F(1, 66) = 0, p < 1.0$]). While there was a significant difference in Uncorrelated Naming Condition RT between experiments (Correlated Naming RT [$F(1, 71) = 0.37, p < 0.54$]; Uncorrelated Naming RT [$F(1, 66) = 50.92, p < 0.001$]), the replication of accuracy scores between experiments allows for direct comparison of individual differences measures.)

Chapter 5: Individual Differences Comparisons, Experiments 1 and 3

Having equated accuracy performance on both the phonological-visual and phonological-orthographic experiments, it is possible to move forward with individual differences comparisons.

Predictions

We expect performance in the Correlated Naming Conditions to be faster and more accurate than in the Uncorrelated Naming Conditions. The following are general predictions for individual differences tasks results with respect to task performance:

Reading Span (Verbal working memory)

Verbal working memory skills should be engaged throughout all of the tasks, as they each require holding verbal auditory information throughout the decision-making process. As such, RS performance may be more predictive as task difficulty increases (that is, as the degree of binding between item names and features decreases).

TOWRE PDE (Phonological decoding efficiency)

Because this is an artificial lexicon experiment, all of the lexical stimuli are necessarily pseudo-words. As with RS, TOWRE PDE may be more predictive as task difficulty increases. Moreover, as the TOWRE PDE task uses common English bigrams to construct its stimuli, predictive-ness of well-bound items could reflect participants' access to, or utilization of, sub-lexical skills.

TOWRE SWE (Word reading efficiency)

Efficient word reading necessitates ease of access to word-level information. If a particular word has regular sound-letter mappings, then the triangle model of word reading suggests that orthographic activation directly activates a word's phonology. For less-bound sound-feature mappings, using just orthography-sound mappings is unhelpful; thus, if TOWRE SWE performance is more predictive of well-bound item learning, that could reflect word learning efficiency.

Face recognition

While not apparently lexical, faces do have share a commonality with words, in that they are familiar visual items comprised of categories of unique parts. Therefore, for our participants, they are word-like both in concept and expertise-level, and face recognition scores could be expected to relate to reading measures, and performance on well-bound items.

Object recognition

Compared to faces, items in the object recognition task are more visually diffuse, and, as a category, are less likely to break into unique components. Object recognition scores are not expected to correlate with many task scores, but perhaps serve as an index of visual difficulty for our task items.

Spoken Word recognition

Since all of the experiments involve the same auditorily presented stimuli, Spoken Word recognition scores are expected to just as an index of task difficulty. If they are predictive, we only expect them to predict task scores for poorly-bound items.

Results

Accuracy and Reaction Time correlations

In the first stage of individual differences analyses, we compared participants' averaged accuracy or RT scores from the experimental task to their scores on each individual differences task (TOWRE SWE, TOWRE PDE, Reading Span, and the Face, Object, and Spoken Word recognition tasks). We will begin with a description of correlations among the individual differences tasks between the two groups for each experiment. Then, for each experiment, we will report significant correlations by Naming Condition. Focus will be on those individual differences correlations that were significant for at least two blocks within the experiment (or within a single block, as well as averaged across the experiment).

Within groups.

Experiment 1. Tables 1 and 2 show the full results of correlations among individual differences measures, by Naming Condition. In the Correlated Naming Condition, TOWRE PDE scores positively and significantly correlated with RS, SWE, and Spoken Word recognition d'

scores ($p < 0.01$ for RS; $p < 0.05$ for SWE and Word d'). Spoken word recognition d' also positively correlated with Object recognition d' ($p < 0.05$). Expectedly, recognition RT scores all correlated with one another ($p < 0.01$).

In the Uncorrelated Naming Condition, relationships amongst the individual differences tasks were quite different, with increased collinearity between tasks. Table 2 shows full correlation results. Unlike in the Correlated condition, RS and SWE performance was highly correlated ($p < 0.001$). Once again, PDE performance correlated with RS and SWE performance ($p < 0.01$); it also positively correlated with Face recognition RT ($p < 0.01$), Object recognition RT ($p < 0.05$), and Spoken Word recognition RT, ($p < 0.001$). Recognition task d' scores and RT scores all positively correlated with one another ($p < 0.001$); additionally, Face recognition RT performance positively correlated with the d' scores from the other recognition tasks ($p < 0.05$).

Experiment 3. Tables 3 and 4 show the full correlation results for the individual differences measures, separated by Naming Condition. Unfortunately, due to an equipment malfunction, some participants are missing TOWRE data (7 in the Correlated Naming Condition, and 5 in the Uncorrelated Naming Condition). While these participants are not included in the TOWRE analyses, they are included in the other individual differences analyses.

For the Correlated Naming condition, unlike in previous groups, accuracy scores for RS, TOWRE SWE, or TOWRE PDE did not significantly correlate. TOWRE SWE scores did significantly (negatively) correlate with RT scores for both Face and Spoken Word recognition ($p < 0.05$), while TOWRE PDE scores positively correlated with Object recognition d' ($p < 0.05$). Object and Spoken Word RT positively correlated with measures of Face recognition d' ($p < 0.05$ for Object RT, $p < 0.01$ for Spoken Word RT). Spoken Word d' positively correlated with Object recognition d' ($p < 0.05$).

Table 1. Individual Differences Correlations Within Experiment Groups, Correlated Naming Condition, Experiment 1.

[illegible]

Table 2. Individual Differences Correlations Within Experiment Groups, Uncorrelated Naming Condition, Experiment 1.

[illegible]

Relationships between the RT tasks followed those from previous experiments: Face recognition RT correlated with both Object and Spoken Word RT ($p < 0.001$); however, Spoken Word and Object RT scores only approached significance ($p < 0.10$).

In the Uncorrelated Condition, RS scores were negatively correlated with measures of Face recognition d' ($p < 0.01$). TOWRE SWE and PDE were correlated with one another ($p < 0.001$); TOWRE SWE scores negatively correlated with Object recognition RT scores ($p < 0.05$), and TOWRE PDE negatively correlated with Face recognition accuracy and d' scores ($p < 0.05$). Face d' scores positively correlated with both scores of Object recognition d' ($p < 0.01$) as well as Spoken Word RT ($p < 0.01$). Additionally, Object recognition d' positively correlated with Spoken Word d' ($p < 0.05$).

There were high instances of collinearity amongst Face recognition task measures ($p < 0.001$ for each) Spoken Word recognition task measures ($p < 0.01$), and Object recognition scores ($p < 0.01$). All measures of recognition RT also positively correlated with one another ($p < 0.001$).

Correlated Naming Condition vs. task accuracy and RT.

Experiment 1. Table 5 describes full results for the Correlated Naming Condition. When examining just this Condition, performance on the Reading Span task was predictive of task accuracy (Blocks 1 and 2, $p < 0.01$; Blocks 3 and 4, $p < 0.05$; averaged across Blocks, $p < 0.01$). TOWRE PDE scores also positively correlated with task accuracy for the first half of the experiment (Blocks 1-3, $p < 0.05$), and averaged across trials ($p < 0.05$). Additionally, face recognition RT performance correlated with RT performance in the Correlated Naming Condition in all Blocks (Block 1, $p < 0.01$; all others, $p < 0.05$), as well as when averaged across trials ($p < 0.05$).

Table 5. Accuracy and Reaction Time Correlations, Experiment 1, Correlated Naming Condition.

Accuracy									
				<u>Face Recognition</u>		<u>Object Recognition</u>		<u>Word Recognition</u>	
<u>Block</u>	<u>RS</u>	<u>SWE</u>	<u>PDE</u>	<u>RT</u>	<u>d'</u>	<u>RT</u>	<u>d'</u>	<u>RT</u>	<u>d'</u>
1	0.45**	0.08	0.34*	-0.12	0.21	0.21	-0.10	0.13	0.05
2	0.44**	0.00	0.31*	-0.29+	0.19	0.20	0.04	0.03	-0.03
3	0.31*	0.09	0.36*	-0.01	0.10	0.18	-0.07	0.07	0.05
4	0.34*	-0.07	0.19	-0.08	0.34*	0.16	0.13	-0.11	0.04
5	0.22	0.08	0.30+	-0.13	0.06	0.11	-0.09	-0.07	-0.01
AVG	0.48**	0.05	0.38*	-0.17	0.23	0.23	-0.04	0.06	0.03
Reaction Time									
				<u>Face Recognition</u>		<u>Object Recognition</u>		<u>Word Recognition</u>	
<u>Block</u>	<u>RS</u>	<u>SWE</u>	<u>PDE</u>	<u>RT</u>	<u>d'</u>	<u>RT</u>	<u>d'</u>	<u>RT</u>	<u>d'</u>
1	-0.10	-0.21	0.10	0.42**	0.04	0.15	0.05	0.14	0.12
2	0.10	-0.21	0.23	0.35*	0.14	0.29+	0.00	0.23	0.19
3	0.07	-0.20	0.23	0.39*	0.15	0.27+	0.00	0.21	0.19
4	0.08	-0.20	0.23	0.38*	0.15	0.26+	-0.01	0.20	0.20
5	0.08	-0.19	0.25	0.38*	0.15	0.27+	-0.01	0.23	0.18
AVG	0.04	-0.20	0.21	0.39*	0.13	0.25	0.00	0.21	0.18

Note: + $p \leq 0.1$, * $p \leq 0.05$, ** $p \leq 0.01$.

78

Accuracy									
				<u>Face Recognition</u>		<u>Object Recognition</u>		<u>Word Recognition</u>	
<u>Block</u>	<u>RS</u>	<u>SWE</u>	<u>PDE</u>	<u>RT</u>	<u>d'</u>	<u>RT</u>	<u>d'</u>	<u>RT</u>	<u>d'</u>
1	-0.16	0.22	0.06	0.12	-0.14	0.14	0.05	-0.19	-0.13
2	0.02	0.51**	0.15	-0.35+	-0.29	-0.30	0.24	-0.32	-0.06
3	0.16	0.47*	0.10	-0.39+	-0.42*	-0.42*	0.24	-0.43*	-0.17
4	0.14	0.41*	0.14	-0.32	-0.35+	-0.37+	0.15	-0.31	-0.19
5	0.07	0.39+	-0.05	-0.39+	-0.34+	-0.43*	0.20	-0.32	0.07
AVG	0.02	0.46*	0.10	-0.27	-0.34+	-0.27	0.20	-0.35+	-0.11
Reaction Time									
				<u>Face Recognition</u>		<u>Object Recognition</u>		<u>Word Recognition</u>	
<u>Block</u>	<u>RS</u>	<u>SWE</u>	<u>PDE</u>	<u>RT</u>	<u>d'</u>	<u>RT</u>	<u>d'</u>	<u>RT</u>	<u>d'</u>
1	-0.22	0.12	0.13	0.23	0.36+	0.46*	-0.13	0.37+	0.08
2	-0.02	0.12	0.12	0.30	0.23	0.17	0.04	0.36+	0.28
3	-0.09	0.01	0.04	0.37+	0.30	0.19	0.06	0.35+	0.41*
4	-0.15	0.10	0.16	0.30	0.28	0.11	0.25	0.33	0.46*
5	-0.37+	0.27	0.12	0.09	0.25	0.13	0.26	0.17	0.43*
AVG	-0.22	0.16	0.14	0.31	0.33	0.29	0.09	0.36+	0.34+
Note: + $p \leq 0.1$, * $p \leq 0.05$, ** $p \leq 0.01$.									

No measurements of recognition task sensitivity significantly correlated with task accuracy or RT in more than one experimental Block.

Experiment 3. Full correlation results are in Table 6. In the Correlated Naming Condition of Experiment 3, task accuracy in the middle Blocks of the experiment was positively correlated with TOWRE SWE performance (Block 2, $p < 0.01$; Block 3, $p < 0.05$; Block 4, $p < 0.05$) and when averaged across trials ($p < 0.05$). Object RT also correlated with task accuracy, albeit negatively (Block 3; $p < 0.05$; Block 5; $p < 0.05$). Spoken Word recognition sensitivity positively correlated with task RT in Blocks 3, 4, and 5 ($p < 0.05$).

Uncorrelated Naming Condition vs. task accuracy and RT.

Experiment 1. Table 7 shows full Correlation results. In the Uncorrelated Naming Condition, TOWRE PDE performance correlated with task accuracy (Block 3, $p < 0.05$; Blocks 4 and 5, $p < 0.01$; averaged across all trials, $p < 0.01$).

Face recognition RT performance also correlated with task RT in the Uncorrelated Naming Condition (Block 1, $p < 0.001$; Blocks 2 and 4, $p < 0.05$; Blocks 3 and 5, $p < 0.01$; and averaged across trials, $p < 0.01$). Object recognition RT correlated positively with task RT in Blocks 1, 2, and 5, and averaged across trials ($p < 0.05$).

While none of the sensitivity indices significantly correlated with task accuracy, Spoken Word recognition sensitivity positively correlated with task RT in all Blocks (Block 1, $p < 0.001$; Blocks 2 - 5, $p < 0.01$) and when averaged across all trials ($p < 0.001$).

Experiment 3. Results for the Uncorrelated Naming Condition are in Table 8. Task accuracy positively correlated with Reading Span performance (Block 2, $p < 0.01$; Block 4, $p < 0.05$; averaged across trials, $p < 0.05$). Task RT positively correlated with measures of

81

Accuracy									
				<u>Face Recognition</u>		<u>Object Recognition</u>		<u>Word Recognition</u>	
<u>Block</u>	<u>RS</u>	<u>SWE</u>	<u>PDE</u>	<u>RT</u>	<u>d'</u>	<u>RT</u>	<u>d'</u>	<u>RT</u>	<u>d'</u>
1	0.27	-0.06	-0.07	0.27	-0.10	0.15	-0.06	0.10	0.32+
2	0.49**	0.07	-0.10	0.03	-0.25	0.02	-0.06	-0.05	0.18
3	0.34+	0.14	-0.06	0.36+	0.05	0.14	0.17	0.29	0.29
4	0.38*	0.23	0.02	0.29	0.02	0.02	0.14	0.21	0.27
5	0.31	0.26	0.03	0.35+	0.06	0.09	0.11	0.26	0.26
AVG	0.42*	0.18	-0.03	0.31	-0.03	0.09	0.09	0.20	0.30

Reaction Time									
				<u>Face Recognition</u>		<u>Object Recognition</u>		<u>Word Recognition</u>	
<u>Block</u>	<u>RS</u>	<u>SWE</u>	<u>PDE</u>	<u>RT</u>	<u>d'</u>	<u>RT</u>	<u>d'</u>	<u>RT</u>	<u>d'</u>
1	0.36+	-0.19	-0.18	0.29	0.05	0.30	0.00	-0.19	0.18
2	0.22	-0.15	-0.05	0.42*	0.19	0.40*	-0.09	-0.15	0.05
3	0.21	-0.17	-0.04	0.33+	0.05	0.26	-0.10	-0.17	0.04
4	0.04	-0.15	-0.08	0.42*	0.19	0.25	0.05	-0.15	0.08
5	0.21	0.04	0.00	0.21	0.08	0.09	0.04	0.04	0.00
AVG	0.22	-0.16	-0.08	0.37+	0.12	0.29	-0.02	-0.16	0.08

Note: + $p \leq 0.1$, * $p \leq 0.05$, ** $p \leq 0.01$.

recognition RT: Face recognition in Blocks 2 and 4 ($p < 0.05$), and Spoken Word RT in Block 1 ($p < 0.05$), Block 2 ($p < 0.01$), Block 4 ($p < 0.05$), and averaged across all trials ($p < 0.05$).

None of the measurements of recognition sensitivity significantly correlated with task accuracy or RT.

Accuracy and Reaction Time Regressions

We used regression analyses to compare individual differences data between averaged accuracy and RT performance in Experiments 1 and 3. A subset of individual differences scores were chosen to minimize collinearity; to achieve a more normal distribution, TOWRE SWE and PDE data were subjected to a logarithmic transform.

Table 9 shows the results of simultaneous regressions of individual differences scores on participants' averaged accuracy scores. There was a significant effect of Condition ($p < 0.001$; Adjusted $R^2 = 0.64$, model $p < 0.001$), and RS ($p < 0.01$); TOWRE PDE was marginally significant ($p < 0.10$). Importantly, there was no significant effect of Experiment-- meaning that our manipulations were effective, and we can more confidently assert that in the case of these two learning experiments, the best predictors of task accuracy were individuals' verbal working memory and phonological decoding skills.

In addition to simultaneous regressions, we used stepwise regression to construct a minimal regression model (Table 10). Compared to the maximal model, this minimal model was only slightly improved in fit (Adjusted $R^2 = 0.66$), and the same predictors as in the simultaneous regression were significant (they were, in fact, the only predictors: Condition, $p < 0.001$; RS, $p < 0.01$; and PDE, $p < 0.10$). An ANOVA of the two regression models showed that the difference between the maximal and minimal models was not significant ($F(115, 124) = 0.44$, $p = 0.91$).

Table 9. Results of Simultaneous Regressions of ID Measures on Task Accuracy, Experiments 1 and 3.

	<u>R^2</u>	<u>B</u>	<u>$SE\ B$</u>	<u>β</u>	<u>P</u>
	0.64				< 0.001 ***
(Intercept)		1.71	0.61		< 0.01 **
Experiment		-0.23	0.18	-0.10	0.22
Condition		-2.01	0.19	-0.88	< 0.001 ***
RS		1.76	0.67	0.15	< 0.01 **
SWE		0.03	0.08	0.02	0.75
PDE		0.18	0.10	0.12	0.08+
Face d'		0.08	0.16	0.03	0.63
Object d'		-0.11	0.18	-0.04	0.53
Spoken Word d'		-0.05	0.19	-0.02	0.78
Face RT		0.00	0.00	0.10	0.23
Object RT		0.00	0.00	-0.01	0.88
Word RT		0.00	0.00	-0.06	0.44
Experiment:Condition		0.27	0.25	0.12	0.27

Table 10. Results of Stepwise Regressions of ID Measures on Task Accuracy, Experiments 1 and 3.

	<u>R^2</u>	<u>B</u>	<u>$SE\ B$</u>	<u>β</u>	<u>P</u>
	0.66				< 0.001 ***
(Intercept)		1.49	0.38		< 0.001 ***
Condition		-1.84	0.12	-0.80	< 0.001 ***
RS		1.79	0.64	0.15	< 0.01 **
PDE		0.16	0.09	0.10	0.07+

Table 11. Results of Simultaneous Regressions of ID Measures on Task RT, Experiments 1 and 3.

	<u>R²</u>	<u>B</u>	<u>SE B</u>	<u>β</u>	<u>P</u>
	0.45				< 0.001***
(Intercept)		6.76	0.18		< 0.001***
Experiment		0.02	0.05	0.03	0.73
Condition		0.43	0.06	0.78	< 0.001***
RS		-0.10	0.20	-0.04	0.62
SWE		-0.01	0.03	-0.02	0.80
PDE		0.03	0.03	0.07	0.38
Face d'		0.03	0.05	0.04	0.57
Object d'		-0.08	0.05	-0.12	0.16
Spoken Word d'		0.13	0.06	0.20	< 0.05*
Face RT		0.00	0.00	0.13	0.22
Object RT		0.00	0.00	0.07	0.44
Word RT		0.00	0.00	0.12	0.26
Experiment:Condition		-0.46	0.07	-0.82	< 0.001***

Table 12. Results of Stepwise Regressions of ID Measures on Task RT, Experiments 1 and 3.

	<u>R²</u>	<u>B</u>	<u>SE B</u>	<u>β</u>	<u>P</u>
	0.46				< 0.001***
(Intercept)		6.70	0.13		< 0.001***
Experiment		0.03	0.05	0.05	0.59
Condition		0.43	0.06	0.78	< 0.001***
Object d'		-0.07	0.05	-0.11	0.15
Spoken Word d'		0.15	0.05	0.23	< 0.01**
Face RT		0.00	0.00	0.15	0.12
Word RT		0.00	0.00	0.16	0.09+
Experiment:Condition		-0.45	0.07	-0.81	< 0.001***

Table 11 shows the results of simultaneous regressions of individual differences scores on averaged task RT for Experiments 1 and 3. Compared to the accuracy model, overall fit for RT was much lower (Adjusted $R^2 = 0.45$, $p < 0.001$). As with Accuracy results, Condition was significant ($p < 0.001$); additionally, there was a significant interaction between Experiment and Condition ($p < 0.001$). Amongst the individual differences tasks, Spoken Word sensitivity was the only significant predictor ($p < 0.05$).

As with task accuracy, we also performed a stepwise regression analysis to construct a minimal regression model (Table 12). Model fit was only slightly better than the full simultaneous regression model (Adjusted $R^2 = 0.46$, $p < 0.001$). As with the full model, there were significant effects of Condition ($p < 0.001$), Spoken Word sensitivity ($p < 0.01$), and Condition-Experiment interaction ($p < 0.001$); Spoken Word RT, in this model, approached significance ($p < 0.10$). An ANOVA of the two regression models showed that the difference between the maximal and minimal models was not significant [$F(115, 120) = 0.39$, $p = 0.86$].

Based on these analyses, we can conclude that both verbal working memory and phonological decoding skill play roles in predicting how accurately participants map novel sounds to novel visual items, regardless of whether those items are strictly orthographic in nature, and one's sensitivity to spoken words may influence how quickly they are able to do so.

General Discussion

Now that we have successfully equated accuracy performance in both the phonological-visual item and phonological-orthographic versions of the experiments, I will address the questions posed initially, when we began our line of questioning about these kinds of learning.

What kinds of individual differences in cognitive abilities might lead to noisy or weak representations/low lexical quality?

Based on my experiments' correlational data, results are diverse. In terms of directly linguistic abilities, phonological decoding skill, verbal working memory, sight word reading, and Spoken Word recognition all played differing roles in predicting novel word learning performance. Additionally, nonlinguistic skills (face and object recognition) played roles. Language is a multimodal domain: concretely, it involves visual (orthographic) and auditory elements; it also includes more abstract elements (grammar, semantics). Our experiment results also suggest that in addition to these established elements of language, other, less obviously-related skill sets (i.e., face and object recognition memory) appear to share influence in linguistic-like learning. However, given the age and experience of the participants, whether the relationships between face and object recognition memory and word learning are causal or coincidental is difficult to parse.

Is performance in learning new words more strongly associated with simple learning (recognition memory) across domains?

When we directly compared experiment accuracy and RT with individual differences, we found that accuracy performance in both learning conditions was significantly predicted by both verbal working memory and phonological decoding skills. That these two skills should be the most predictive of task performance is particularly interesting for Experiment 1, where no actual orthography was used. In the absence of an actual orthographic system, participants were likely using the features of the visual stimuli as a de facto orthography (at least in the Correlated Naming condition), assigning phonology to the pictures themselves.

That Reading Span is significant along with TOWRE PDE scores bears some mentioning. The Reading Span task, while ostensibly a measure of verbal working memory, necessitates other skills for successful completion (syntactic processing, word reading, etc.). In

this way, it is the most complex task in our individual differences task battery. While it is a standard measure for studies of individual differences and word reading, recent research suggests that due to its complexity, RS may be only spuriously related to reading comprehension; that is, relationships between RS and experimental task scores may not be purely due to shared verbal working memory skills, but to the other elements of lexical quality that are necessary to perform the RS task, such as word reading (Van Dyke, Johns, & Kukona, 2014).

Given this argument, the pairing of both phonological decoding and verbal working memory as significant predictors of performance in a novel word learning task makes sense. But before verbal working memory is discounted entirely, it is worth noting that Van Dyke and colleagues (2014) point out that, in the case of experiments like ours, the case for RS as a measure of verbal working memory is much stronger. More caution should be exercised for studies where sentence reading skill is the dependent factor: sentence reading at the expert level is automatic, and with the exception of the times where re-evaluation is necessary, very little working memory ability is necessary. However, for our tasks, where the holding of phonological information is necessary while the participant makes a decision, verbal working memory is necessary.

Stepping back to our original questions about language learning: the experiments here suggest that language learning (and, by association, difficulty) is the result of a diverse set of 'systems' acting in concert. Weakness in one, or strength in another, can cascade between modalities, affecting learning outcomes. So in this way, these experiments appear not only to support the tenets of the LQH, but also to extend them.

Chapter 6: Experiment 4

Introduction

The final experiment in this dissertation extends the paradigms of the preceding experiments, combining phonological-visual learning, phonological-orthographic learning, and coherent covariation of linguistic features of novel items in a within-subjects design. Importantly, the level of typicality between an item and its features (that is, the degree of binding between constituents of an artificial lexicon) is manipulated in more detail than in previous experiments. In the previous experiments, items were either be perfectly well-bound, or perfectly poorly-bound, between a visual stimulus and its name; in this final experiment, items' features vary between categories, so that items range from being perfectly typical on all levels for a category to perfectly atypical on all levels. In the following experiment, I explore how these varying levels of constituent binding affect novel word learning, and take the first steps towards integrating the concepts of constituent binding (from the LQH) and coherent covariation (from

semantic cognition/concepts and categories); additionally, I examine how individual lexical quality interacts with word learning and predicts performance in the task.

Methods

Participants

Twenty-four University of Connecticut undergraduate students participated in the experiment; all were monolingual American English speakers with normal hearing and normal or corrected-to-normal vision.

Apparatus and Materials

For this experiment, I designed 12 novel line-drawn items belonging to two different animal categories (fish or butterflies; see Figure 38 for an example), differing in degree of "coherence" between lexical dimensions: phonology (the first syllable of the name) or semantics (the color of the item). Items were either predictable in their dimensions (by having the most typical feature for their animal category within the artificial lexicon), or unpredictable in one or both dimensions (see Table 13 for a complete description of stimuli design). Additionally, each of the 12 novel items had a unique and corresponding pair of novel orthographic stimuli. This allowed me to take preliminary steps towards addressing (a) whether increasing typicality facilitates performance (due to coherence with category structure) or impairs it (due to competition among items overlapping in features), (b) whether there are any differences due to overlap in phonological vs. semantic features, and (c) whether high LQH/constituent binding/coherence promotes learning orthography.







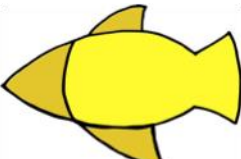



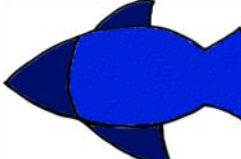

Image	Orthography	Phonology
		“Pi-Gau”
		“Pi-Wu”
		“Pi-Lou”
		“Gu-Pae”
		“Gu-Bi”
		“Pi-Tei”

Figure 38. Example Stimuli for Experiment 4.

As in the previous experiments, phonological stimuli were recorded in a sound-attenuated booth with a sampling rate of 44.1 kHz. They were then edited and checked for artifacts and clipping using Praat software. Visual stimuli were created by scanning hand-drawn line drawings, which were then edited using InkScape software.

Procedure

The experiment took place in several stages:

Individual differences measures. As in my previous experiments, TOWRE SWE, TOWRE PDE, and Reading Span were administered, to measure single word reading efficiency, sublexical decoding efficiency, and verbal working memory, respectively. Additionally, I added a measurement of reading comprehension ability (Nelson-Denny Reading Test), and a measure of IQ equivalence (WASI-2, Weschler Abbreviated Scale of Intelligence, matrix reasoning). During the Nelson-Denny reading comprehension test, participants were presented with a packet of eight passages of prose, followed by four to eight questions related to each passage (for a total of 36 questions). Participants were told to read through each passage and answer each multiple-choice question following the passage, not lingering too long on any single question. They were given 15 minutes to answer as many questions as correctly as possible. The WASI-2 test of matrix reasoning is a measure of general IQ, and consists of 25 multiple-choice questions. Participants were shown 25 unique picture arrays, and told to choose one of 5 items that best completed the pattern or array at the top of the page. They were graded for accuracy as well as time required to complete the test.

Exposure Phase 1 (Phonological and Semantic). Participants were told that they were to learn the names of newly discovered animals. First, they were shown the items onscreen, paired with simultaneous auditory presentation of the item's name (binaurally, via Sennheiser

HD-595 headphones). Participants were shown each item twice (for a total of 24 exposures), and allowed to self-progress through this Phase. This was an innovation not included in previous studies; based on Trudeau's unpublished dissertation (2006), we expected this to boost initial performance above chance in the following Phase.

Training/Test Phase 1 (Phonological and Semantic). As in my previous experiments, each trial began when the participant clicked on a cross in the center of the screen; reaction time recording was initiated at the mouse click. Participants were immediately shown two visual items on a computer screen and simultaneously asked to click on the item they thought was being named (as in previous experiments, participants were presented with the carrier sentence, e.g., "Find /piwu/"). After each choice, participants were given feedback on whether the choice was the correct one. Trials did not end until the correct item was chosen; if a participant chose the incorrect item, they were asked to "Try again" until they chose the correct item. When the participant clicked on the correct item, the incorrect item disappeared, and they heard feedback like "That's right, that's /piwu/," and then the trial ended (the entire feedback period was 2300 ms after clicking on the Target). A 1000 ms blank-screen inter-trial interval followed; to begin the next trial, participants clicked on a cross in the center of the computer screen. Participants continued this Phase by Block (132 trials/Block) until a criterion of 84% was reached (for most participants, it took two Blocks, or 264 trials, to reach criterion). Reaction time, and accuracy per trial were measured.

Exposure Phase 2 (adding orthography). Participants were then shown a novel orthography: as in the first exposure Phase, participants were presented the items (visually and auditorally) alongside a novel orthographic name (the novel orthography was the same type of orthography used in the previously mentioned experiments; kept separate for ease of learning).

Also as in the first exposure Phase, participants were allowed to self-progress, for a total of 24 exposures (two exposures per item).

I will note that in this experiment, unlike my previous experiments utilizing an artificial orthography, there was no coherence manipulation for the orthography-phonology pairings; orthographic representations consisted of a left-side character that corresponded to the first syllable, and a right-side character that corresponded to the second syllable of the item name. As in Experiment 3, orthographic characters were clearly separate. Thus, participants were exposed to 14 individual characters: 2 first-syllable characters, and 12 unique second-syllable characters.

Training/Test Phase 2. A similar 2AFC paradigm as in Training/Test Phase 1 was used, though this time with the addition of orthography. For each trial, the Target item was presented at the top center of the screen (i.e., the upper portion of the screen), along with its auditory name. Underneath, two choices of orthography were presented, and participants were asked to click on the one they thought was being named. As in the first testing Phase, feedback was given, and trials progressed until the correct item was chosen; participants repeated this Phase by Block (132 trials/Block) until a criterion of 84% was reached (typically, only one run of the Block was necessary). Reaction time and accuracy per trial were taken in the same manner as in previous experiments.

Results

Accuracy and RT

Phase and Block. As this was a learning task to criterion, participants' accuracy performance improved as the task progressed (Figures 39-42 illustrate accuracy and RT progression throughout the experiment). RT was slower in Phase 2; this could be the result of increased deliberation, or a side effect of having an additional item onscreen during the trial.

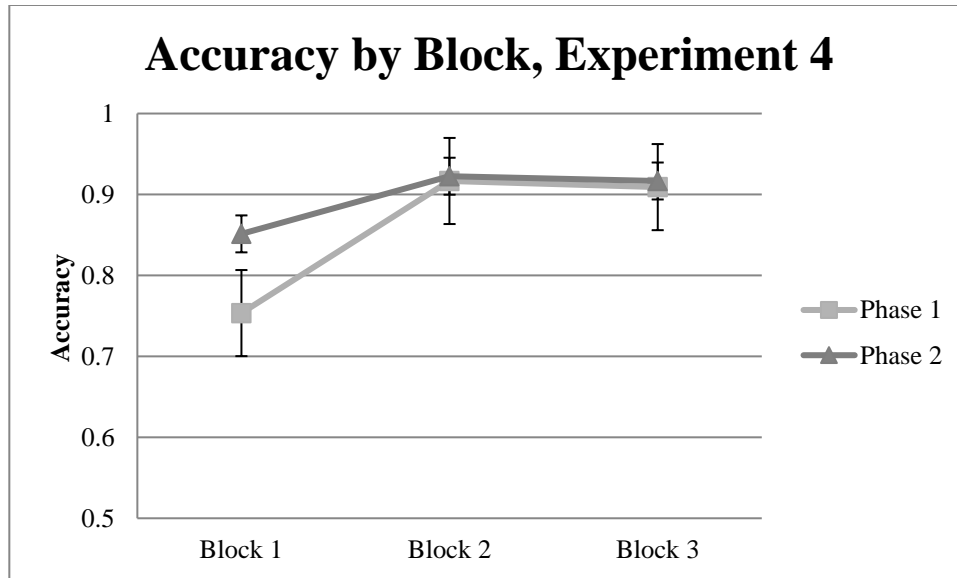


Figure 39. Accuracy scores for Experiment 4, separated by Phase and Block. Each Block was 132 trials; participants could not progress to the next Phase until they reached an accuracy criterion of 84% correct trials. Number of participants per Block, Phase 1: $n_{\text{Block } 1}=24$; $n_{\text{Block } 2}=19$; $n_{\text{Block } 3}=3$. Number of participants per Block, Phase 2: $n_{\text{Block } 1}=24$; $n_{\text{Block } 2}=9$; $n_{\text{Block } 3}=1$.

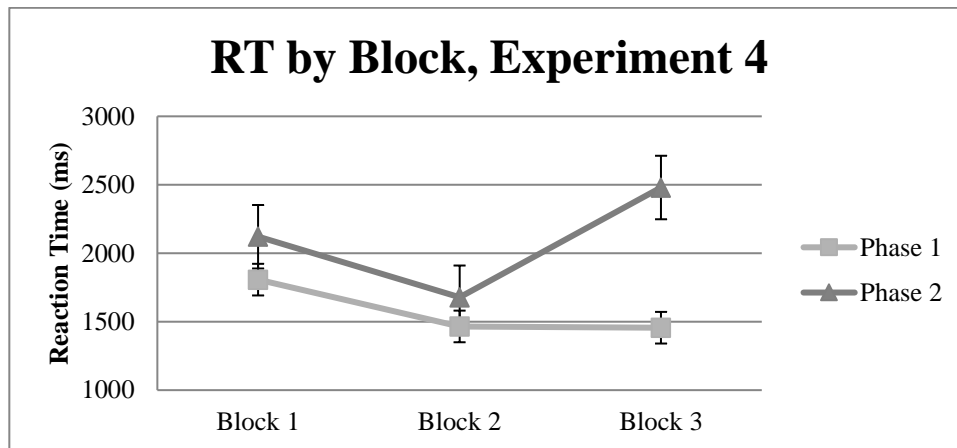


Figure 40. RT scores for Experiment 4, separated by Phase and Block. Each Block was 132 trials; participants could not progress to the next Phase until they reached an accuracy criterion of 84% correct trials. Number of participants per Block, Phase 1: $n_{\text{Block } 1}=24$; $n_{\text{Block } 2}=19$; $n_{\text{Block } 3}=3$. Number of participants per Block, Phase 2: $n_{\text{Block } 1}=24$; $n_{\text{Block } 2}=9$; $n_{\text{Block } 3}=1$.

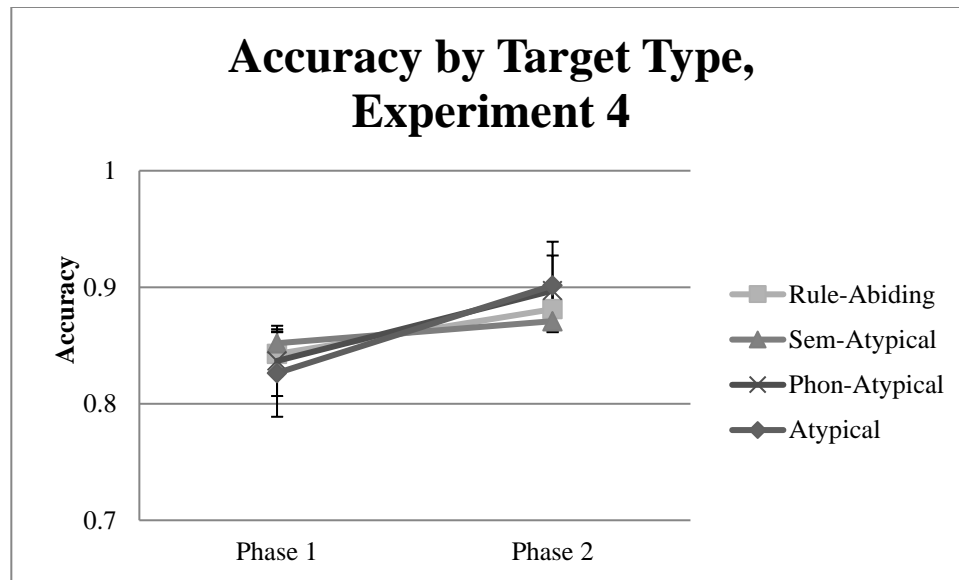


Figure 41. Accuracy scores for Experiment 4, separated by Phase and Target Type.

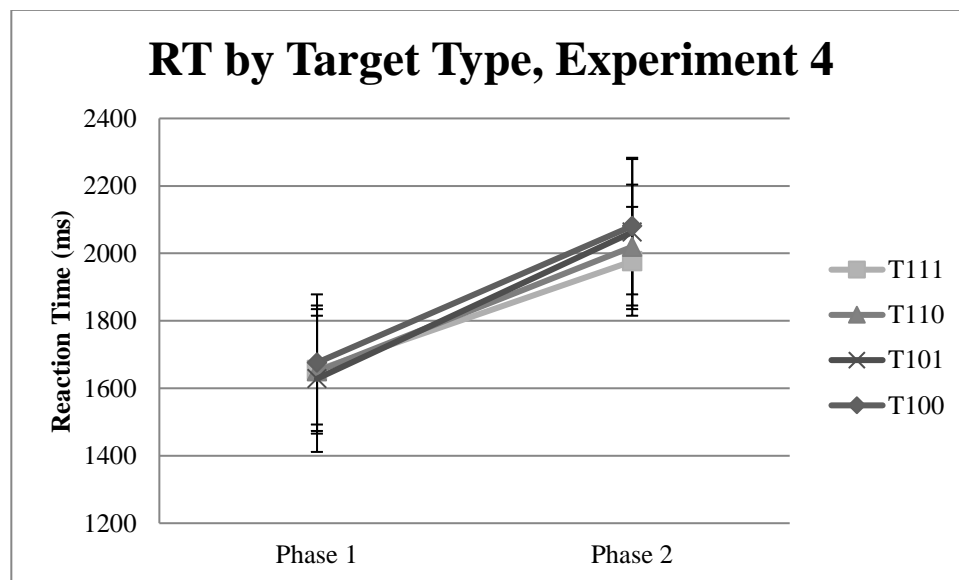


Figure 42. RT scores for Experiment 4, separated by Phase and Target Type.

Analyses of variance on accuracy scores did not show a main effect of Phase [$F(1, 22) = 0.12$, $p < 1.0$], though they did show a main effect of Block [Mean accuracy in Block 1 = 0.80; Block

2= 0.92; and Block 3= 0.92; $F(1, 21) = 6.49, p < 0.01$]. Given that task Phase progression relied on criterion accuracy by Block, this is to be expected. There was no significant interaction on accuracy between Phase and Block. Follow-up Block-level accuracy ANOVAs showed that there was a significant effect of Phase for Block 1 [$F(1, 2759) = 68.36, p < 0.001$], but not Blocks 2 [$F(1, 1602) = 2.51, p < 1.0$] or 3 [$F(1, 2) = 0.29, p < 1.0$; Figure 39].

When examining reaction time, analyses of variance did not show main effects of either Phase [$F(1, 22) = 1.20, p < 1.0$] or Block [$F(2, 21) = 0.86, p < 1.0$]. While Figure 40 does show visual differences between Phases in Block 3, this large (but statistically insignificant) difference is likely due to the small number of participants in each Block ($n = 3$ in Phase 1, and $n = 1$ in Phase 2).

Given that the accuracy scores between Phases are only significant for the first Block (and subsequent differences in RT for later Blocks are likely driven by the comparatively smaller numbers of participants), further analyses of Target Type and Lexical Competition effects will only focus on the first Block of each Phase.

Target Type. Neither accuracy scores in Phase 1 nor Phase 2 showed a significant main effect of Target Type in ANOVA analyses (Phase 1 [$F(3, 1365) = 2.582, p < 0.10$]; Phase 2 [$F(3, 1365) = 1.57, p < 1.0$]). This was also true for reaction time (Phase 1 [$F(3, 1158) = 0.74, p < 1.0$]; Phase 2 [$F(3, 1256) = 0.71, p < 1.0$]).

Lexical Competition Type, by Phase. Finally, analyses of variance were performed for each type of Target-Competitor overlap: Category Match (whether items shared the same semantic category), Phonological Match (whether items shared the same first syllable), or Semantic Match (whether items shared the same color). These analyses of variance were performed with two levels of Trial Type (Match, or Mismatch), and 1 level of Block

Category Match. While accuracy scores in Phase 1 did not show a significant effect of Category Match on accuracy [$F(1, 1367) = 0.39, p < 1.0$] they did in Phase 2 (Mean Match= 0.83; Mean Mismatch= 0.87; [$F(1, 1367) = 8.31, p < 0.01$]). That this difference appears in the second Phase is likely due to participants' learning of category associations in the first Phase, therefore creating scaffolding upon which to add information in the orthography phase—and thus creating opportunities for lexical competition, as participants were significantly more accurate for Category Mismatch trials.

There was no significant effect of Category Match on RT scores in either Phase (Phase 1 [$F(1, 1160) = 0.64, p < 1.0$]; Phase 2 [$F(1, 1258) = 0.82, p < 1.0$]).

Phonological Match. Accuracy scores in Phase 1 did not show a significant effect of Phonological Match on accuracy [$F(1, 1367) = 0.36, p < 1.0$]. However, in Phase 2, there was a significant difference (Mean Match= 0.81; Mean Mismatch= 0.89; [$F(1, 1367) = 24.44, p < 0.001$]), as Phonological Mismatch trials were more accurate than cohort trials. Likewise, for RT, while scores in Phase 1 did not show a significant effect of Phonological Match on accuracy [$F(1, 1160) = 1.39, p < 1.0$], they did in Phase 2 (Mean Match= 2435 ms; Mean Mismatch= 1946 ms; [$F(1, 1258) = 55.34, p < 0.001$]). This is a clear case of lexical competition in the second Phase—in the cohort trials, accuracy was lower, and reaction times slower, in trials that had overlapping syllables.

Semantic Match. Neither accuracy scores in Phase 1 nor Phase 2 showed a significant effect of Semantic (Color) Match on accuracy in an ANOVA analyses (Phase 1 [$F(3, 1365) = 2.582, p < 0.10$]; Phase 2 [$F(3, 1365) = 1.57, p < 1.0$]). The same was true for RT scores (Phase 1 [$F(1, 1160) = 0.22, p < 1.0$]; Phase 2 [$F(1, 1258) = 0.001, p < 1.0$]).

Individual Differences Measures

Within participants. Table 14 shows individual differences correlation results amongst participants. While RS performance positively correlated with TOWRE PDE performance ($p < 0.05$), TOWRE SWE and PDE scores did not correlate. TOWRE PDE scores did correlate with Nelson Denny raw ($p < 0.05$), but not composite, scores (raw scores are not normed to grade level, as composite scores are—as we only used one part of the Nelson-Denny test, raw scores were kept in data analysis).

Table 14. Individual Differences Correlations Within Participants, Experiment 4.

	<u>RS</u>	<u>SWE</u>	<u>PDE</u>	<u>WASI-2 (Raw)</u>	<u>ND (Raw)</u>
<u>SWE</u>	0.35				
<u>PDE</u>	0.43*	0.23			
<u>WASI-2 (Raw)</u>	0.04	0.28	0.07		
<u>ND (Raw)</u>	0.38+	0.01	0.43*	0.13	
<u>ND (Comp)</u>	0.11	-0.40+	0.19	-0.33	0.30
Note: + $p \leq 0.1$; * $p \leq 0.05$; ND=Nelson Denny Comprehension.					

Accuracy and Reaction time correlations. Initial individual differences analyses were performed by comparing participants' averaged accuracy or RT scores from the experimental task to their scores on each individual differences task (Reading Span, TOWRE SWE, TOWRE PDE, Nelson-Denny comprehension, and WASI-2 Matrix reasoning). Significant results are first reported by Phase; within Phase, they are then reported by Block, and then by Target type within the Phase. Tables 15 and 16 contain full descriptions of the correlation results.

Phase 1. On average, accuracy in Phase 1 (regardless of number of trials) was positively correlated with TOWRE sight word reading ($p < 0.05$); TOWRE phonological decoding approached significance ($p < 0.10$). Additionally, time spent in Phase 1 (as measured by number of trials) was negatively correlated with performance on TOWRE phonological decoding

($p < 0.01$), suggesting that the better participants were at phonological decoding, the faster they reached criterion in the initial Phase of the experiment.

When separated by Block, accuracy performance in Phase 1 positively correlated with TOWRE PDE in the first Block ($p < 0.05$), and Reading Span ($p < 0.01$) and TOWRE PDE in the second Block ($p < 0.01$; $n = 19$). Only 3 participants needed a third Phase, so this data was exempt from individual differences analyses. Task RT correlations in Phase 1 only showed significant correlations in the second Block (TOWRE SWE, $p < 0.05$, and Nelson Denny raw scores, negative, $p < 0.05$), suggesting that word reading and comprehension ability negatively predicted time spent during correct trials (that is, those with better word reading and more correct comprehension scores had shorter trial times).

When separated by Target type (averaged across trials), accuracy for Rule-Abiding items was positively correlated with TOWRE performance (SWE, $p < 0.05$). For the Wholly Atypical items, Reading Span scores ($p < 0.05$) and TOWRE performance (SWE, $p < 0.01$; PDE, $p < 0.05$) predicted accuracy performance, suggesting that learning these Atypical items' names most taxed participants' verbal working memory abilities and word-reading skills. Additionally, Nelson Denny composite scores approached significance (negatively, $p < 0.1$) with regard to accuracy scores, and Nelson Denny composite scores approached significance for RT ($p < 0.1$). Performance on trials where the Target was atypically phonological, but semantically typical, did not correlate with any individual differences measures.

Phase 2. In the second Phase, averaged across all trials, both pseudoword decoding (TOWRE PDE, $p < 0.01$) and comprehension scores (Nelson Denny raw and composite scores, $p < 0.05$)

correlated positively with task accuracy. While word-level skills (TOWRE PDE and SWE) were predictive of performance in Phase 1, the addition of comprehension performance may reflect acquisition of a deeper level of lexical quality for the novel items—something that can only be done in the second phase. When examining number of trials in Phase 2, Nelson Denny composite scores negatively predicted task performance ($p<0.05$); that is, better comprehenders learned the orthography faster. Also, TOWRE PDE performance approached significance (also negatively, $p<0.10$).

In the first Block of the second Phase, as in the average trials, performance on the TOWRE PDE and Nelson Denny comprehension tasks predicted task accuracy (TOWRE PDE, $p<0.01$; Nelson Denny raw scores, $p<0.05$; Nelson Denny composite, $p<0.01$). In Block 2, ($n=9$), Nelson Denny composite scores best predicted task accuracy ($p<0.05$). As in the first Phase, fewer than 4 participants needed a third Block to complete the experiment, so the third Block was excluded from individual differences correlations.

When separated by Target Type (averaged across Blocks), accuracy for Rule-Abiding items was positively correlated with TOWRE PDE and Nelson Denny composite scores ($p<0.05$); Nelson Denny raw scores approached significance ($p<0.1$). For Phonologically Atypical items, TOWRE SWE performance predicted task accuracy ($p<0.01$). For Semantically Atypical items, Nelson Denny raw scores approached significance in predicting task accuracy ($p<0.1$). Finally, for Wholly Atypical items, TOWRE SWE performance positively predicted task accuracy ($p<0.01$), while TOWRE PDE scores approached significance ($p<0.1$).

As this is a task of language learning, it is unsurprising that the language-specific tasks are the best predictors of task performance. However, it is worth noting that phonological decoding skill is the most prevalent predictor, in both Phases of the task. To harken back to the

first chapter of the dissertation, this would most support the assertions of the Phonological Deficit Hypothesis (and by extension, the LQH—and suggest that if one tenet of lexical quality is most crucial, it may be phonology).

Number of trials. Comparing task performance at its most basic (that is, how many trials it took to reach criterion; Table 17) shows that for the phonological-visual learning, phonological decoding was the best predictor, and comprehension came into play for phonological-orthographic learning. This shift of predictors could be indicative of the increased scaffolding available to the learners—perhaps the relationship would be different in a task like that of Experiment 3.

Simultaneous Regressions. As in previous experiments, individual differences measures and task accuracy were then subjected to full regression analyses, and then compared to stepwise regression analyses to construct a minimal regression models. Unfortunately, the adjusted R-squared values for each of these models was extremely low, accounting for less than 3% of the variance. In the future, a larger sample size may help to avoid such a result.

Discussion

Typicality and Word Learning

This learning experiment's results paint an interesting picture of how individual items' features may interact with participants' performance. In the first Phase of the experiment, participants were simply matching names to pictures. In the second Phase, participants built upon a foundation of sound-picture pairings, adding a novel orthography. Participants were more accurate at trials for items between categories (i.e., trials with both a fish and a butterfly), and were both less accurate and slower when the two items shared the same first syllable. Together, this suggests that the phonological and semantic features of the lexicon are critical components

of task performance. This is a particularly interesting result to have found in the second Phase: participants have all learned the phonological-semantic connections, but phonological and category competition between items is still significantly affecting task performance.

Individual Differences and Lexical Quality

Generally, correlations showed that word-level tasks (TOWRE SWE, TOWRE PDE) were predictive of accuracy performance in the first Phase of the experiment. Reading Span and Nelson-Denny Comprehension had a trade-off: the former was predictive of performance in Phase 1, while the latter was predictive of performance in Phase 2. This could reflect the change in task demand: in Phase 1, word learning was simply phonological-visual, which would necessitate working memory; in Phase 2, the addition of an orthographic component, and explicit reference to the non-orthographic visual form of the Target, likely recruited similar mechanisms to reading comprehension.

When examining correlations by Target type, TOWRE SWE significantly predicted accuracy on typical targets in Phase 1, while Reading Span, TOWRE SWE, and TOWRE PDE performance was correlated with accuracy for Wholly Atypical items. That Reading Span is significantly predictive of accuracy for Atypical items mirrors analogous findings from previous experiments: Reading Span significantly predicted accuracy performance in the Uncorrelated Naming Condition in Experiment 3. Neither Phonologically nor Semantically Atypical items had significant predictors in the first Phase.

In Phase 2, TOWRE PDE performance was predictive of accuracy on typical Target trials, as were Nelson-Denny Composite scores (this was the only Target type for which comprehension was a significant predictor). For Wholly Atypical items, TOWRE SWE was the only significant predictor; it also predicted accuracy performance on Phonologically Atypical

trials. At first pass, the relative predictability of the TOWRE tasks in this orthographic learning task seems reversed: phonological decoding ability is predictive of performance on Typical items, and sight word efficiency is most predictive for Atypical items. (The explanation for this finding may bring us back to the triangle model: for Typical items, one still may need to rely on the breakdown of the parts (as in TOWRE PDE)—after all, knowing that an item is a typical blue fish still leaves the participant with three choices as to the correct answer. For Atypical items, participants may be learning their features more holistically, since breaking down the constituent parts (as one must do when using TOWRE PDE) is not necessarily helpful, compared to the Typical item trials.)

Given that nearly all of the predictors uniquely predicted some aspect of task performance, it is difficult to definitively choose just one element of paralinguistic ability that is most “crucial” to novel word learning. Initially, this was my aim in creating these experiments; combined with my previous chapters, it seems that the opposite claim is more apparent: when creating new lexical representations, it is all-skills-on-deck. The stages in which these skills manifest themselves may differ, but all seem to be similarly integral to lexicon building.

Chapter 7: General Discussion

This dissertation has taken a nuanced approach towards investigating a very complex activity; but given the number of mental operations required to successfully process language—and consequently, the myriad of theorized ways in which this system could fail, or be perturbed—such an approach is required. By breaking down the process of word learning into its constituent parts, we were able to investigate how one's individual strengths in some aspects of linguistic and paralinguistic ability affect the incorporation of new words into the lexicon, and what this may mean for existing theories of reading ability/the mental lexicon.

In the first three experiments, I created and expanded a paradigm for examining the relationship between skilled readers' existing memory and linguistic abilities and their abilities to learn novel lexical items that differed diametrically in terms of sound-feature relatedness. In my final experiment, I combined the elements of the previous experiments (artificial lexicons using

orthographic, phonological, and visual items, and individual differences measures) into a single experiment, and furthered the sound-feature relatedness manipulations of the novel items by creating a continuum of "typicality" of an item for its given category.

This work was all done with the Lexical Quality Hypothesis (Perfetti, 2007; Perfetti & Hart, 2002) as the guiding element. As outlined in the first chapter, the LQH accommodates many aspects of language in its assertions concerning why a reader may comprehend the way they do. This makes it very flexible, and as such, it can account for many theoretical stances and empirical findings (as briefly discussed in the first chapter). However—and this is not meant to be a critique but rather, an observation—the LQH lacks specificity, a hierarchy that other theories of reading and comprehension ability hinge upon. In performing the experiments described in this dissertation, I sought to answer if, within the elements of lexical quality outlined by Perfetti and Hart, certain elements of lexical quality were more crucial than others.

My individual differences measures formed a small battery of language and memory-related assessments: the Reading Span task for a measure of verbal working memory, TOWRE tests of Phonological Decoding and Sight Word Efficiency, the Nelson Denny test of Reading Comprehension, WASI-2 test of Matrix reasoning, and tests of Face, Object, and Spoken Word recognition. All of these tests significantly predicted task performance in some form. An overarching aim in this dissertation was to find some hierarchical specificity within lexical quality—but after four different experiments, it is difficult to declare a clear victor.

While our artificial lexicon experiments were designed to control for as much external variability in learning for our participants as possible, it is clear that language learning does not occur in a vacuum. At least in these experienced, typical college readers, novel word learning involves—or rather, may be bolstered by—many aspects of reading or memory ability, even

when such skills are not apparently called for (as in the case of face or object recognition with orthography learning).

Thus, our examination of the Lexical Quality Hypothesis vis-à-vis novel word learning has shown that if anything, it is even more accommodating than initially suggested, as paralinguistic capabilities can also predict lexical learning as well.

Some caveats should be considered, however, before such a conclusion can be settled: while the previous work was definitely rooted in, and inspired by, theories of word reading, the experimental paradigm did not directly involve reading. While the LQH is concerned explicitly with reading comprehension, it is still a cognitive model, and therefore I would argue that such transference between language modalities is warranted. In fact, focus on cognitive and paralinguistic skills might be necessary, if a researcher wishes to tease apart individual differences in high-level readers.

While this shift of focus from strictly lexical to lexical and cognitive skills may seem like an odd one, other studies not strictly focused on reading comprehension have taken advantage of the flexibility afforded by the LQH in attempts to either expand upon the definitions of quality or to explore how lexical quality—either within the reader, or between lexical items—interacts with other skills.

For example, Gilbert, Goodwin, Compton, & Kearns (2013) examined the utility of morphological awareness skills in predicting reading comprehension in developing (fifth-grade) readers. Using multisyllabic word reading as a measure of morphological awareness, they found that its relationship with reading comprehension was most pronounced with poor readers (and not pronounced in better readers. This finding suggests that while it may be an integral aspect of lexical quality, morphological awareness' effect on reading comprehension may be minimized

by the relative strength of other skills. Compared to poor readers', Gilbert et al. (2014) describe better readers' lexical representations as "encapsulated," that is, the individual components are more efficiently or automatically accessed during reading. In other words, the more tightly bound representations of better readers made it more difficult to measure any single component as integral to word reading skill. This may speak to the diffuse conclusion this dissertation work, as all of the participants, while they were on a spectrum of lexical or memory ability, were still all expert or high-achieving readers, simply by virtue of being university undergraduates.

Considering these high-achieving readers in terms of their long-term reading development may be helpful. Longitudinal studies of young readers (from elementary through middle school) have documented the relative shift of individual skills' predictability of comprehension skill. For example, Verhoeven and Van Leeuwe (2008) examined over 2000 Dutch schoolchildren from ages 6-12, approximately. While vocabulary skills were consistently predictive of reading comprehension skill as the students progressed, students' word decoding skills became less predictive of students' reading comprehension abilities. This is not to say that word-decoding skill was no longer a significant predictor of reading comprehension-- after all, we found a significant correlation between the two skill-sets in our own undergraduate population-- but its level of predictability did diminish. While the slope of reading skill is much steeper for developing readers than for skilled readers, but I still would argue that changes in reading are present across the lifespan-- and that, when considering indicators of lexical quality in highly skilled readers, some creativity in choosing cognitive domains may be warranted.

In fact, relationships between lexical quality and non-obviously linguistic domains have been found. Veldre & Andrews (2014) found evidence for relationships between eye movements and participants' lexical quality. When tracking participants' eye movements during a moving-

window sentence reading task (that is, during reading, participants were only shown a few of the characters surrounding their point of fixation at a given time), they found that those with higher lexical representations (as indexed by tests of reading comprehension and spelling accuracy) both had greater saccade lengths and performed more poorly in subsequent comprehension tasks when parafoveal (around the point of gaze fixation) information was removed. While it is not likely (or at least, not argued by either Veldre & Andrews or myself) that non-disordered eye movements alone contribute to lexical quality, the idea that something so pluripotent as eye movements can be reflexive of individual lexical quality speaks to the idea that recruitment of extra-linguistic skills can also be indices of lexical quality.

Even in conjunction with these previous studies, there is still more exploration of word learning and lexical quality to be had. While the experiments in this dissertation explored phonology, orthography, and coherence between the two in several ways, the semantic elements of lexical quality were comparatively shallow. If one were to extend this paradigm, a logical next step would be to add some additional, deeper semantic characteristics to the novel items (it is possible that the two familiar categories of animal were not different enough for that task). While the idea of semantics being an important part of lexical quality is non-controversial, a deeper and more nuanced manipulation of semantic features in novel items could provide more insight and clarity into which para-linguistic skills may be most recruited when incorporating new items into the lexicon.

Additionally, prosody is an important feature of spoken language that not only helps to enrich semantic representations of words (National Assessment of Educational Progress, 1995), but is also absent from descriptions of lexical quality (Perfetti, 2002; Perfetti and Hart, 2007). The use of prosodic, or expressive, tones when reading aloud imparts additional information

about a word's meaning to the listener (Rasinski, 2004), and as such, should be an element of lexical quality. While the auditory stimuli in this dissertation were created to control for prosodic differences, future research focused on exploring the bounds of lexical quality could incorporate such information into a novel word learning experiment.

References

- Amitay, S., Ben-Yehudah, G., Banai, K., Ahissar, M. (2002). Disabled readers suffer from visual and auditory impairments but not from a specific magnocellular deficit. *Brain*, 125, 2272-2285.
- Castles, A., & Coltheart, M. (1993). Varieties of developmental dyslexia. *Cognition*, 413, 149–180.
- Cohen, L., Jobert, A., LeBihan, D., & Dehaene, S. (2004). Distinct unimodal and multimodal regions for word processing in the left temporal cortex. *Neuroimage*, 23, 1256-1270.
- Cohen, L., Lehericy, S., Chochon, F., Lemer, C., Rivaud, S., & Dehaene, S. (2002). Language-specific tuning of visual cortex? Functional properties of the Visual Word Form Area. *Brain*, 125, 1054-1069.
- Cornelissen, P., Bradley, L., Fowler, M.S., & Stein, J.F. (1992). Covering one eye affects how some children read. *Developmental Medicine & Child Neurology*, 34, 296–304.
- Cornelissen, P.L, Richardson, A.R., Mason, A., Fowler, M.S., & Stein, J.F. (1994). Contrast sensitivity and coherent motion detection measured at photopic luminance levels in dyslexics & controls. *Vision Research*, 315(10), 1483–1494.
- Cutting, L. E., Clements-Stephens, A., Pugh, K. R., Burns, S., Cao, A., Pekar, J. J., Davis, N., & Rimrodt, S. L. (2013). Not all reading disabilities are dyslexia: Distinct neurobiology of specific comprehension deficits. *Brain Connectivity*, 3(2), 199-211.
- Daneman, M., & Carpenter, P.A. (1980). Individual differences in working memory and reading. *Journal of Verbal Learning and Verbal Behavior*, 19, 450–466.

- Eden, G. F., VanMeter, J. W., Rumsey, J. W., Maisog, J. and Zeffiro, T. A. (1996). Functional MRI reveals differences in visual motion processing in individuals with dyslexia. *Nature*, 382, 66–69.
- Everatt, J. (1999). *Reading and Dyslexia—Visual and Attentional Processes*. Routledge: London.
- Facoetti, A., Lorusso, M. L., Cattaneo, C., Galli, R., & Molteni, M. (2005). Visual and auditory attentional capture are both sluggish in children with developmental dyslexia. *Acta Neurobiologiae Experimentalis*, 65, 61-72.
- Facoetti, A., Lorusso, M. L., Paganoni, P., Cattaneo, C., Galli, R., & Mascetti, G. G. (2003). The time course of attentional focusing in dyslexic and normally reading children. *Brain and Cognition*, 53, 181–184.
- Facoetti, A., Paganoni, P., Turatto, M., Marzola, V., & Mascetti, G. G. (2000). Visuospatial attention in developmental dyslexia. *Cortex*, 36, 109–123.
- Facoetti, A., Turatto, M., Lorusso, M. L. & Mascetti, G. G. (2001). Orienting of visual attention in dyslexia: evidence for asymmetric hemispheric control of attention. *Experimental Brain Research*, 138, 46-53.
- Galaburda, A. M., & Livingstone, M. (1993). Evidence for a magnocellular defect in developmental dyslexia. In P. Tallal, A.M. Galaburda, R.R. Llinas and C. von Euler (Eds), *Temporal Information Processing in the Nervous System. Annals of the New York Academy of Science*, 682, 70–82.
- Galaburda, A. M., Menard, M. T., and Rosen, G. D. (1994). Evidence for aberrant auditory anatomy in developmental dyslexia. *Proceedings of the National Academy of Science USA*, 913, 8010–8013.

- Gilbert, J. K., Goodwin, A. P., Compton, D. L., & Kearns, D. M. (2014). Multisyllabic word reading as a moderator of morphological awareness and reading comprehension. *Journal of Learning Disabilities, 47*(1), 34-43.
- Hari, R., Valta, M., & Uutella, K. (1999). Prolonged attentional dwell time in dyslexic adults. *Neuroscience Letters, 271*, 202–204.
- Harm, M. W., & Seidenberg, M. S. (1999). Reading acquisition, phonology, and dyslexia: Insights from a connectionist model. *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.
- Hart, L., & Perfetti, C. A. (2008). Learning words in Zekkish: Implications for understanding lexical representation. In E. L. Grigorenko & A. J. Naples (Eds.), *Single word reading: Behavioral and biological perspectives*, (107-128). New York: Taylor & Francis.
- Hartley, D. E. H., & Moore, D. R. (2002). Auditory processing efficiency deficits in children with developmental language impairments. *Journal of the Acoustical Society of America, 112*, 2962–2966.
- Hickok, G., & Poeppel, D. (2004). Dorsal and ventral streams: A framework for understanding aspects of the functional anatomy of language. *Cognition, 92*, 67–99.
- Iles, J., Walsh, V., & Richardson, A. (2000). Visual search in dyslexia. *Dyslexia: An International Journal of Research and Practice, 6*(3), 163–177.
- Livingstone, M. S., Rosen, G. D., Drislane, F. W., & Galaburda, A. M. (1991). Physiological and anatomical evidence for a magnocellular defect in developmental dyslexia. *Proceedings of the National Academy of Science USA, 88*, 7943–7947.

- Lovegrove, W. J., Martin, F., Blackwood, M. and Badcock, D. (1980). Specific reading difficulty: Differences in contrast sensitivity as a function of spatial frequency. *Science*, 210, 439–440.
- Luce, P. A., & Pisoni, D. B. (1998). Recognizing Spoken Words: The neighborhood activation model. *Ear and hearing*, 19(1), 1-36.
- Magnuson, J. S., Kukona, A., Braze, D., Johns, C. L., Van Dyke, J. A., Tabor, W., Mencl, W.E., Pugh, K.R., & Shankweiler, D.P. (2010). Phonological instability in young adult poor readers. *Proceedings of the 32nd annual conference of the Cognitive Science Society*, 1429-1434.
- Magnuson, J.S., Tanenhaus, M. K., Aslin, R.N., & Dahan, D. (2003). The time course of Spoken Word learning and recognition: Studies with artificial lexicons. *Journal of Experimental Psychology: General*, 132, 202-227.
- McAnally, K.I., & Stein, J.F. (1996). Abnormal auditory transient brainstem function in dyslexia. *Proceedings of the Royal Society B.*, 263, 961–965.
- McClelland, J. L., & Elman, J. L. (1986). The TRACE model of speech perception. *Cognitive Psychology*, 18(1), 1-86.
- McClelland, J. L., & Rogers, T. T. (2003). The parallel distributed processing approach to semantic cognition. *Nature*, 4, 310-322.
- Medin, D. L., & Schaffer, M. M. (1978). Context theory of classification learning. *Psychological Review*, 85(3), 207.
- Menell, P., McAnally, K.I., & Stein, J.F. (1999). Psychophysical and physiological responses to amplitude modulations in dyslexia. *Journal of Speech & Hearing Research*, 42, 797–803.

- Moore, B.C.J. (1989). *An Introduction to the Psychology of Hearing, 3rd Edition*. Academic Press: London.
- Murphy, G. L., & Medin, D. L. (1985). The role of theories in conceptual coherence. *Psychological Review*, 92(3), 289-316.
- Nagarajan, S., Mahncke, H., Salz, T., Tallal, P., Roberts, T., & Merzenich, M. M. (1999). Cortical auditory signal processing in poor readers. *Proceedings of the National Academy of Sciences of the United States of America*, 96(11), 6483-6488.
- National Assessment of Educational Progress (NAEP), 1995. Listening to Children Read Aloud: Oral Fluency. <<http://www.nces.ed.gov/pubs95/web/95762.asp>> (retrieved 08.25.14).
- Nestor, A., & Tarr, M. J. (2008). Gender recognition of human faces using color. *Psych Sci*, 19(12), 1242-1246.
- Olson, R.K., Wise, B., Conners, F., Rack, J., & Fulker, D. (1989). Specific deficits in component reading and language skills—genetic and environmental influences. *Journal of Learning Disabilities*, 22, 339–348.
- Perfetti, C. (2007). Reading ability: Lexical quality to comprehension. *Scientific Studies of Reading*, 11(4), 357-383.
- Perfetti, C.A., & Hart, L. (2002). The lexical quality hypothesis. In L. Verhoeven, C. Elbro & P. Reitsma (Eds.), *Precursors of functional literacy* (Vol. 11). Amsterdam/ Philadelphia: John Benjamins Publishing.
- 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., & Devlin, J. T. (2003). The myth of the visual word form area. *NeuroImage*, 19, 473–481.
- Rasinski, T.V. (2004). *Assessing reading fluency*. Pacific Resources for Education and Learning: Honolulu, Hawai'i.
- Renvall, H., & Hari, R. (2003). Diminished auditory mismatch fields in dyslexic adults. *Annals of Neurology*, 53(5), 551-557.
- Rosch, E. (1975). Cognitive representations of semantic categories. *Journal of Experimental Psychology: General*, 104(3), 192-233.
- Rosch, E., & Mervis, C. B. (1975). Family resemblances: Studies in the internal structure of categories. *Cognitive Psychology*, 7(4), 573-605.
- Rumelhart, D. E., & McClelland, J. L. (1981). Interactive processing through spreading activation. *Interactive Processes in Reading*, 37-60.
- Sabatini, J. P., Sawaki, Y., Shore, J. R., & Scarborough, H. S. (2010). Relationships among reading skills of adults with low literacy. *Journal of Learning Disabilities*, 43(2), 122–138.
- Schulte-Körne, G., Deimel, W., Bartling, J., & Remschmidt, H. (1998). Auditory processing and dyslexia: evidence for a specific speech processing deficit. *Neuroreport*, 9, 337–340.
- Schulte-Körne, G., Deimel, W., Bartling, J., & Remschmidt, H. (2001). Speech perception deficit in dyslexic adults as measured by mismatch negativity (MMN). *International Journal of Psychophysiology*, 40, 77–87.
- Seidenberg, M. S., & McClelland, J. L. (1989). A distributed, developmental model of word recognition and naming. *Psychological Review*, 96, 523-568.

- Shankweiler, D., & Crain, S. (1986). Language mechanisms and reading disorder: A modular approach. *Cognition*, 24(1–2), 139–168.
- Shaywitz, B.A., Shaywitz, S.E., Blachman, B.A., Pugh, K.R., Fulbright, R.K., Skudlarski, P., Mencl, W.E., Constable, R.T., Holahan, J.M., Marchione, K.E., Fletcher, J.M., Lyon, G.R., & Gore, J.C. (2004). Development of left occipitotemporal systems for skilled reading in children after a phonologically-based intervention. *Biological Psychiatry*, 55(9), 926-933.
- Smith, E. E., & Medin, D. L. (1981). *Categories and Concepts* (51-83). Cambridge, MA: Harvard University Press.
- Sperling, A. J., Lu, Z.-L., Manis, F. R., & Seidenberg, M. S. (2005). Deficits in perceptual noise exclusion in developmental dyslexia. *Nature Neuroscience*, 9, 862–863.
- Sperling, A.J., Lu, Z.-L., Manis, F.R., & Seidenberg, M.S. (2006). Motion-perception deficits and reading impairment. *Psychological Science*, 17(12), 1047–1053.
- Stein, J. (2001). The magnocellular theory of developmental dyslexia. *Dyslexia*, 7(1), 12-36.
- Stein, J.F. & McAnally, K.I. (1996). Impaired auditory temporal processing in dyslexics. *Irish Journal of Psychology*, 16, 220–228.
- Stein, J.F., & Walsh, V. (1997). To see but not to read; the magnocellular theory of dyslexia. *Trends in Neuroscience*, 20, 147–152.
- Talcott, J.B., Hansen, P.C., Elikem, L.A., & Stein, J.F. (2000). Visual motion sensitivity in dyslexia: evidence for temporal and motion energy integration deficits. *Neuropsychologia*, 38, 935-943.

- Talcott, J.B., Hansen, P.C., Willis-Owen, C., McKinnell, I.W., Richardson, A.J., & Stein, J.F. (1998). Visual magnocellular impairment in adult developmental dyslexics. *Neuroophthalmology*, 20, 187–201.
- Talcott, J.B., Witton, C., McClean, M., Hansen, P.C., Rees, A., Green, G.G.R., & Stein, J.F. (1999). Can sensitivity to auditory frequency modulation predict children's phonological and reading skills? *Neuroreport*, 10, 2045–2050.
- Talcott, J.B., Witton, C., McClean, M., Hansen, P.C., Rees, A., Green, G.G.R., & Stein, J.F. (2000). Visual and auditory transient sensitivity determines word decoding skills. *Proceedings of the US National Academy of Science*, 97, 2952–2958.
- Torgesen, J. T., Wagner, R. K., & Rashotte, C. A. (1999). Test of word reading efficiency (TOWRE). Austin, TX: PRO-ED.
- Trudeau, J. (2006). *Meaningful Pseudowords Demonstrate Semantic Contributions to Single-Word Naming*. (Unpublished doctoral dissertation). University of Connecticut, Storrs, CT.
- Van den Noort, M., Bosch, P., Haverkort, M., & Hugdahl, K. (2008). A standard computerized version of the reading span test in different languages. *European Journal of Psychological Assessment*, 24(1), 35-42.
- Van Dyke, J. A., Johns, C. L., & Kukona, A. (2014). Low working memory capacity is only spuriously related to poor reading comprehension. *Cognition*, 131, 373-403.
- Veldre, A., & Andrews, S. (2014). Lexical quality and eye movements: Individual differences in the perceptual span of skilled adult readers. *The Quarterly Journal of Experimental Psychology*, 67(4), 703-727.

- Verhoeven, L., & Van Leeuwe, J. (2008). Prediction of the development of reading comprehension: a longitudinal study. *Applied Cognitive Psychology*, 22, 407-423.
- Witton, C., Richardson, A., Griffiths, T.D., Rees, A., & Green, G.D.R. (1997). Temporal pattern analysis in dyslexia. *British Journal of Audiology*, 31, 100–101.
- Witton, C., Talcott, J. B., Hansen, P. C., Richardson, A. J., Griffiths, T. D., Rees, A., Stein, J. F., & Green, G. G. (1998). Sensitivity to dynamic auditory and visual stimuli predicts nonword reading ability in both dyslexic and normal readers. *Current Biology*, 8(14), 791-797.
- Yoncheva, Y. N., Blau, V. C., Maurer, U., & McCandliss, B. D. (2010). Attentional focus during learning impacts N170 ERP responses to an artificial script. *Developmental Neuropsychology*, 35(4), 423-445.