

The impact of frequency, consistency, and semantics on reading aloud:

An artificial orthography learning paradigm



Jo S. H. Taylor

St John's College, University of Oxford

Thesis submitted for the degree of Doctor of Philosophy, Trinity Term, 2009

TABLE OF CONTENTS

Acknowledgements.....	i
Short Abstract.....	ii
Long Abstract.....	iii
Chapter 1: General introduction.....	1
Chapter 2	
Semantic influences on orthographic processing: A novel word learning experiment....	40
Chapter 3	
How do we learn spelling-sound mappings? An artificial orthography paradigm.....	75
Chapter 4	
The influence of proficiency on discrimination and generalization.....	113
Chapter 5	
The role of semantics in learning to read.....	140
Chapter 6	
Is phonological or semantic familiarity important in learning to read.....	188
Chapter 7: General discussion.....	229
References.....	266
Appendices.....	286

ACKNOWLEDGEMENTS

First and foremost I would like to thank my supervisors, Professors Kate Nation and Kim Plunkett. Kate has always been hands-on, guiding me through research design and results presentation and helping me to achieve clarity in my writing. Kim has provided me with a way to step back from my work and consider its wider theoretical implications. I am also extremely grateful to my research group (LCD) and other members of the department for providing me with a friendly atmosphere in which to pilot, conduct, and discuss my work. In particular, Jessie Ricketts has been a constant source of support and an excellent role model. Thanks also goes to my supervisors for encouraging me to apply for the studentship I was awarded from the Economic and Social Research, and Medical Research councils, without which I would not have been able to undertake my doctoral work. For financial support I would also like to acknowledge St. John's College, Oxford. Their generosity both in terms of conference funding and accommodation has greatly enriched my experience as a doctoral student. Through St. John's I also met the friends who have carried me through my DPhil. I would particularly like to mention Rachel Jones and Tom Fenlon for putting up with my moods and always being there to make me smile. For being my working, cooking, and laughing partner, Andrew Wynn deserves a huge thank you. Last but certainly not least I want to thank my loving family. I would not have come to Oxford as an undergraduate, nor stayed on to undertake a DPhil, without their constant encouragement and unceasing belief in me.

SHORT ABSTRACT

The impact of frequency, consistency, and semantics on reading aloud:

An artificial orthography learning paradigm

Jo S. H. Taylor

St John's College, University of Oxford

Thesis submitted for the degree of Doctor of Philosophy, Trinity Term, 2009

Five experiments explored how we learn to read and recognise words with typical and atypical spelling-sound mappings and to generalize to novel words. In Experiment 1, adults learned to read pseudowords with typical or atypical pronunciations. There was some evidence that prior exposure to word meanings enhanced orthographic learning. However, interpretation was clouded by stimulus control problems that plague research using natural alphabets. In Experiment 2, an artificial orthography paradigm was developed to overcome these problems. Adults learned to read novel words written in novel symbols. Post-training, they could generalize, indicating extraction of individual symbol sounds. The frequency and predictability of symbol-sound mappings influenced learning and generalization, mirroring natural language findings. Experiment 3 found extended training to improve item recognition and generalization. In Experiment 4, pre-exposure to item sounds plus an object referent vs. item sounds provided equivalent benefit for orthographic learning. By the end of training, this was limited to items with low frequency unpredictable symbol-sound mappings. In Experiment 5, pre-exposure to novel definitions enhanced orthographic learning more than pre-exposure to item sounds, but by the end of training, both conditions were again equally beneficial.

LONG ABSTRACT

The impact of frequency, consistency, and semantics on reading aloud:

An artificial orthography learning paradigm

Jo S. H. Taylor

St John's College, University of Oxford

Thesis submitted for the degree of Doctor of Philosophy, Trinity Term, 2009

One of the most impressive aspects of the human language faculty is our ability to cope with both rules and exceptions to those rules. This is particularly well illustrated by English orthography. How does a skilled reader of English read words with both typical (e.g., MAT or JUMP) and atypical spelling-sound mappings (e.g., PUT or DONE), whilst maintaining the ability to read novel words such as JAT, LUT, or RONE? The experiments reported in this thesis addressed this question by exposing adults to novel words which varied in the predictability of their spelling-sound mappings and examining learning and generalization.

The triangle model of reading aloud (Plaut, McClelland, Seidenberg, & Patterson, 1996) offers one explanation for how we learn to read. In this model, knowledge of spelling-sound mappings emerges through exposure to the orthographic forms of whole words (and their corresponding pronunciations). This knowledge is sensitive to the context in which letters occur. For example, the model learns that although OO is most commonly pronounced /u/ (e.g., FOOD, SPOON, ROOM), OOK is typically pronounced /ʊk/ (e.g., BOOK, TOOK, LOOK), at least in received pronunciation (RP). It therefore

pronounces a novel word such as VOOK to rhyme with BOOK. The model is also sensitive to word frequency; HAVE is experienced so often that it is learned well, despite being pronounced differently to other words with similar spellings (e.g., GAVE, PAVE, SAVE).

The triangle model further suggests that knowledge of word meanings (semantics) is helpful when learning to read words with atypical spelling-sound mappings, particularly when they are also low in frequency. These words are difficult to learn and semantic support therefore increases the system's efficiency. For example, when learning to read the word WATCH, knowing that it tells the time may help us to remember to pronounce it differently from words such as MATCH, CATCH, and BATCH.

Behavioural research offers some support for the idea that reading aloud and visual word recognition are influenced by spelling-sound frequency and predictability in both adults (Jared, 2002; Jared, McRae, & Seidenberg, 1990; Lacruz & Folk, 2004) and children (Weekes, Castles, & Davies, 2006). There is also evidence that we use context-sensitive spelling-sound knowledge when generalizing to novel words (Treiman, Kessler, & Bick, 2003; Treiman, Kessler, Zevin, Bick, & Davis, 2006). Furthermore, the semantic properties of words have been found to influence orthographic processing. For example, Strain, Patterson, and Seidenberg (1995; 2002) found that if a word was low in frequency and had an atypical spelling-sound mapping, it was read faster if it was high imageability (e.g., DOUGH or SWAMP) than if it was low imageability (e.g., DREAD or SCARCE).

Unfortunately, such natural language research suffers from difficulties in quantifying and measuring lexical variables and has often failed to control for correlated

factors. For example, Monaghan and Ellis (2002) demonstrated that if age-of-acquisition (AoA) is taken into account, imageability no longer influences reading aloud. AoA effects may stem from the fact that early acquired words have stronger phonological representations (Brown & Watson, 1987; Morrison & Ellis, 1995; Morrison, Ellis, & Quinlan, 1992). Thus, imageability effects may in fact reflect a phonological rather than a semantic variable. This highlights the difficulty of separating familiarity with word meanings from familiarity with word sounds. To explain further, earlier it was suggested that knowing the meaning of an inconsistent word, such as WATCH, might help us to pronounce it differently from MATCH, CATCH and BATCH. However, it might simply be that our familiarity with the sound of the word /wɒtʃ/ but not /wætʃ/ helps us to remember how to read this word.

Experiment 1 (presented in Chapter 2) used a novel word learning method to investigate the role of semantic vs. phonological familiarity in orthographic processing. Eighty adults learned to read 24 novel words which could be pronounced by analogy to English words. Each participant learned 12 items which were assigned a consistent pronunciation (e.g., MAVE pronounced to rhyme with GAVE), and 12 items which were assigned an inconsistent pronunciation (e.g., GLORM pronounced to rhyme with WORM). Prior to this, participants learned either high or low imageability definitions of the items (*semantic pre-exposure*), the sounds of the items (*lexical phonology pre-exposure*), or received *no pre-exposure*.

Semantic pre-exposure increased subsequent reading accuracy, both during training and at post-test, relative to the no pre-exposure baseline. This was particularly

evident for inconsistent items. Lexical phonology pre-exposure provided intermediate benefit. These findings offer some support for the idea that familiarity with word meanings supports reading aloud over and above familiarity with word sounds. However, some doubt is cast on this interpretation by the finding that high imageability definitions were not more beneficial than low imageability definitions. This left open the possibility that definitions did not benefit reading aloud due to their semantic content but because learning them increased attention to item sounds. Furthermore, pre-exposure did not have a facilitative effect on the speed with which participants could read items or discriminate trained from untrained items at post-test. This contrasts with research showing that semantic variables impact on naming (Balota, Cortese, Sergent-Marshall, Spieler, & Yap, 2004; Strain et al., 1995) and lexical decision latencies (Balota et al., 2004; Pexman, Haregreaves, Siakaluk, Bodner, & Pope, 2008).

Two major problems were highlighted with the method used in Experiment 1. First, novel words were written in English orthography meaning that proficiency was extremely high from the beginning of training. The experiment therefore failed to address the impact of semantics on *learning* to read. Second, although research suggests that semantic influences on reading aloud vary according to word frequency and spelling-sound typicality, we in fact know very little about how sensitivity to these lexical variables develops. For example, to what extent is spelling-sound knowledge acquired implicitly through experience with words vs. explicitly through teaching rules? The main conclusion drawn from Experiment 1 was that new methods are needed to investigate the role of frequency, consistency, and semantics in orthographic *learning*.

The remaining experiments presented in this thesis addressed these problems using an artificial orthography paradigm in which adults learned to read novel words written in novel characters. This provided complete control over exposure to the statistics of the language and enabled direct examination of the process by which we learn spelling-sound mappings. In Experiment 2 (Chapter 3), 12 adults learned to read 36 novel words written in novel characters. Vowels in the novel words varied in the frequency and consistency of their character-sound mappings. Some vowel characters were entirely *consistent* and had only one pronunciation, while others were *inconsistent* and had two different pronunciations which depended on the consonant onset. In addition, some vowels were highly frequent, occurring in many words, and some were low in frequency, occurring in few words. Participants were not taught individual character sounds, only whole-word pronunciations. Once they had learned to read the novel words, post-tests included discrimination of trained from untrained items (old-new decision) and reading a further set of novel words written in the same characters as the training set (generalization).

Experiment 2 demonstrated that following 30-45 minutes of training, adults could read the novel orthography to at least 80% accuracy. Post-training, they were also able to discriminate trained from untrained items at above chance levels, and read generalization items to 73% accuracy. To achieve such high levels of reading accuracy, particularly in generalization, learners had to develop sensitivity to conditional relationships between consonant onsets and inconsistent vowel pronunciations. These results provide the first conclusive evidence that knowledge of context-sensitive

character-sound mappings can be extracted through exposure to the orthographic form of whole words and their pronunciations, as suggested by the triangle model of reading aloud. Furthermore, the frequency and consistency of character-sound mappings interacted in their effects on learning in a way that mirrored observations from natural languages (Jared, 2002; Weekes et al., 2006), and, again, the behaviour exhibited by the triangle model. Overall, these findings demonstrate that an artificial orthography paradigm provides a high degree of experimental control and generates data that simulate typical reading processes.

Although Experiment 2 demonstrated that participants could discriminate trained from untrained items at above chance levels, this process was extremely laboured and was subject to false positive responding. In addition, discrimination was insensitive to the frequency and consistency characteristics of trained items, in contrast to lexical decision in natural languages (e.g., Lacruz & Folk, 2004). Similarly, frequency and consistency effects were less strong in generalization than in training. Experiment 3 (Chapter 4) investigated whether extended training on the novel orthography would improve discrimination and generalization and increase the influence of frequency and consistency in these tasks. Sixteen adults learned to read the artificial orthography to $\geq 70\%$ accuracy on Day 1 and $\geq 90\%$ accuracy on Day 2. On both days, training and testing proceeded as in Experiment 2 with one modification: old-new decision instructions were clarified to ensure that participants knew that they were required to judge whether they recognised the whole item, not whether the characters were familiar. Old-new decision and generalization were indeed enhanced on Day 2 relative to Day 1. Consistency effects

in old-new decision also emerged on Day 2. However, even on Day 1, old-new decision times were sensitive to frequency, and overall performance on this task was enhanced relative to that seen in Experiment 2. In addition, generalization was influenced by frequency and consistency on both days. Overall, these results suggest that the most important improvements on Experiment 2 resulted not from extended training but from modifications to task instructions and increased participant numbers.

Experiment 4 (Chapter 5) returned to the question of whether semantics plays a role in learning to read. Existing research suggests that semantic factors are more important when reading low frequency inconsistent words. Experiments 2 and 3 demonstrated that frequency and consistency produce strong and reliable effects on learning in the artificial orthography. This means that the paradigm is particularly suitable for addressing the role of semantics. Furthermore, because participants are unfamiliar with the novel orthography before training, the influence of semantics on orthographic *learning* could be directly assessed.

Thirty-two participants learned to read the artificial orthography to $\geq 70\%$ accuracy. Sixteen participants were pre-exposed to the association between item sounds and novel objects (*semantics*) and 16 participants were pre-exposed to item sounds (*lexical phonology*). Relative to a no pre-exposure baseline, provided by Day 1 of Experiment 3, semantic pre-exposure improved reading accuracy during training. By the end of training, this facilitation was restricted to items containing low frequency inconsistent vowels. Semantic pre-exposure also enhanced discrimination of trained from untrained items. These findings seem to support the role for semantics embodied

in the triangle model of reading aloud. However, this is contradicted by the fact that lexical phonology pre-exposure produced equivalent facilitation. Thus, familiarity with word meanings may not benefit orthographic learning beyond the contribution provided by familiarity with word sounds.

A further finding to emerge from Experiment 4 was that generalization accuracy was lower following pre-exposure, relative to the no pre-exposure baseline. As pre-exposure increased participants' familiarity with item sounds, they may have been able to read trained items by recalling that they were a valid word in the language. While this might boost trained item reading, it could reduce the necessity of extracting context-sensitive character-sound relationships. This would be detrimental for generalization.

A caveat to the argument that semantic familiarity does not benefit orthographic learning, over and above familiarity with word sounds, is that novel objects may have induced only weak semantic representations. This possibility was investigated in Experiment 5 (Chapter 6) in which novel definitions rather than objects were used to instantiate semantics. Thirty-two participants learned to read the artificial orthography. Prior to this they were pre-exposed to the sounds of all items (*lexical phonology*) and to novel definitions for half of these items (*semantics*). Semantic pre-exposure enhanced reading accuracy during training relative to lexical phonology pre-exposure. By the end of training, facilitation was specific to items containing low frequency inconsistent vowels. Discrimination of trained from untrained items was also more accurate in the semantic condition. As in Experiment 4, pre-exposure reduced generalization accuracy, relative to the no pre-exposure baseline provided by Day 1 of Experiment 3.

Taken alone these findings provide strong support for the role of semantics in orthographic processing embodied in the triangle model of reading aloud. However, there is an alternative explanation that must be ruled out before this conclusion can be accepted. In Experiment 5, semantic vs. lexical phonology pre-exposure was manipulated within subject; participants were required to learn definitions for semantic items and this demanding task was interspersed with blocks of lexical phonology pre-exposure trials in which they just had to listen to and repeat the items. Thus, an efficient strategy would have been to concentrate on learning the definitions of semantic items and to pay minimal attention during the lexical phonology listen-repeat task. In contrast, in Experiment 4, the between subjects pre-exposure manipulation ensured that participants in the lexical phonology condition could focus their full attention on learning item sounds. It is therefore possible that the benefit of semantic over lexical phonology pre-exposure seen in Experiment 5 was not a genuine semantic effect, and instead resulted from depressed performance in the lexical phonology condition.

This possibility was investigated by comparing performance in Experiment 5 to the lexical phonology condition from Experiment 4, with reference to the no pre-exposure baseline (Day 1, Experiment 3). These analyses demonstrated that semantic pre-exposure in Experiment 5 enhanced reading accuracy in the early and middle stages of orthography training to a greater extent than lexical phonology pre-exposure in Experiment 4. Thus, familiarity with word meanings does benefit the early stages of learning to read aloud, over and above familiarity with word sounds. However, in end of training performance analyses, semantic pre-exposure in Experiment 5 provided

equivalent benefit to lexical phonology pre-exposure in Experiment 4 for items containing low frequency inconsistent vowels. Furthermore, end of training performance on these items was lower in the lexical phonology condition in Experiment 5 than in the equivalent condition from Experiment 4. This pattern of results was to a large extent mirrored in the old-new decision data. These findings suggest that learning definitions did depress performance in the lexical phonology condition. Overall, the results of this experiment indicate that while knowledge of word meanings does benefit orthographic learning, knowledge of word sounds is also important. Previous experiments that have neglected the correlation between these variables may have overestimated the role of semantics in orthographic processing.

A key area for the future will be to examine how the roles of phonological and semantic familiarity in orthographic learning change with increasing experience with a novel orthography. Developmental research also suggests that extended orthography training might eliminate the depressed generalization performance observed following pre-exposure in Experiments 4 and 5 (Landi, Perfetti, Bolger, Dunlap, & Foorman, 2006). Of equal importance will be to investigate whether providing richer semantic knowledge increases the influence of this variable on orthographic learning. Overall, the experiments reported in this thesis demonstrate that artificial orthography paradigms represent an exciting direction for new research on learning to read. They generate data which simulate typical reading processes, provide a high degree of experimental control, and can be used to directly probe orthographic learning.

CHAPTER 1: GENERAL INTRODUCTION

This thesis investigates a central question in research on reading: How do we learn to read and recognise words with both typical and atypical spelling-sound mappings while maintaining the ability to generalize and read novel words? The nature of this problem is clearly illustrated by the following poems:

*I take it you already know
Of tough and bough and cough and dough?
Others may stumble, but not you,
On hiccough, thorough, lough and through?
Well done! And now you wish, perhaps,
To learn of less familiar traps?
Beware of heard, a dreadful word
That looks like beard and sounds like bird,
And dead: it's said like bed, not bead --
For goodness sake don't call it 'deed'!
Watch out for meat and great and threat
(They rhyme with suite and straight and debt).
A moth is not a moth in mother,
Nor both in bother, broth in brother,
And here is not a match for there
Nor dear and fear for bear and pear;
And then there's dose and rose and lose --
Just look them up -- and goose and choose,
And cork and work and card and ward,
And font and front and word and sword,
And do and go and thwart and cart --
Come, come, I've hardly made a start!
A dreadful language? Man alive!
I'd mastered it when I was five!*

(Anonymous poem believed to have first appeared in The Times, 1936)

*Tw'as brillig, and the slithy toves
Did gyre and gimble in the wabe:
All mimsy were the borogoves,
And the mome raths outgrabe.*

(Jabberwocky, Lewis Carroll, 1872)

The nature of the problem

Reading aloud involves translating a word's printed form (orthography) to its spoken form (phonology). In languages with alphabetic writing systems this can be achieved by converting letters to sounds, e.g., CAT = /k - æ - t/, and blending these sounds together, e.g., /kæt/. Decades of research on reading development have revealed much about when these decoding skills develop (Ehri, 1992; 1985), what lower-level skills underlie them (Bradley & Bryant, 1983; Rack, Hulme, Snowling, & Wightman, 1994), and why some children have particular difficulty learning to decode (Marshall, Snowling, & Bailey, 2001; Mody, Studdert-Kennedy, & Brady, 1997; Snowling, Gallagher, & Frith, 2003). However, in English orthography, and to a lesser extent in certain other alphabetic scripts, some words are not easy to read using regular letter-sound rules. For example, in the word WHAT, the pronunciation of A is different to that in CAT, and the H is also silent. This thesis investigates a question that remains poorly understood: how do we learn to read and recognise irregular, or inconsistent, words, such as WHAT, while maintaining the ability to generalize and read novel words like DAT?

The relationship between item-specific and general knowledge is a central question in psycholinguistics (Fodor & Pylyshyn, 1988; Glushko, 1979; Humphreys & Evett, 1985; Pinker, 1991) and has been the focus of a number of modelling initiatives using a variety of different architectures (Coltheart, Curtis, Atkins, & Haller, 1993; Coltheart, Rastle, Perry, Langdon, & Ziegler, 2001; Harm & Seidenberg, 1999;2004; Jacobs & Grainger, 1994; Pinker & Prince, 1988; Plaut, McClelland, Seidenberg, &

Patterson, 1996; Plunkett & Juola, 1999; Rumelhart & McClelland, 1986; Seidenberg & McClelland, 1989; Zorzi, Houghton, & Butterworth, 1998). However, relatively little research has directly examined the learning process by which we gain such abilities. In the experiments reported in this thesis, adults learn to read novel words which vary in the typicality of their spelling-sound patterns. Their ability to read and recognise these words and to generalize their knowledge to novel words is assessed. This enables direct investigation of the factors influencing orthographic learning. In this introductory chapter, existing research on the factors that influence reading aloud and visual word recognition is first reviewed, along with how models of reading account for these data. Within this, effects of word frequency and spelling-sound predictability and the role of word meanings are explored. Following this review, problems with the existing research are delineated and some examples are given of how research in other domains has tried to overcome similar problems. Finally, the questions addressed in this thesis are outlined.

Throughout this thesis, the terms *reading aloud* and *naming* will refer to tasks in which overt pronunciation is required. *Visual word recognition* will describe the process of judging whether the written form of a presented item is a word in the target language (i.e., deciding if it is a word or a nonword). The term *orthographic processing* will be used to encompass both of these outcome measures. Similarly, *orthographic learning* will refer to the development of the ability to read aloud and recognise the visual forms of words. *Generalization* will always describe tasks in which novel orthographic forms

must be overtly pronounced. It is also worth noting that examples adhere to received pronunciation.

The factors that influence orthographic processing

Spelling-sound predictability and word frequency

Coltheart et al. (1993) proposed that we possess grapheme-phoneme correspondence rules (GPCs) which convert letters (or letter sequences) to their most common single phoneme pronunciation. For example, MEAT would be converted to /m/ - /i/ - /t/. Words can then be defined as regular, if all graphemes are pronounced according to GPCs (e.g., DUCK), or irregular/exceptional, if they contain graphemes that are pronounced differently (e.g., DEAF - EA most commonly pronounced to rhyme with LEAF). GPCs enable us to read regular words and to generalize and read novel words. But how do we cope with irregular or exception words? Coltheart et al. suggested that these are stored in a whole-word, or lexical, manner and that these lexical representations are sensitive to the frequency with which an item has been encountered.

The distinction between regular and irregular words is captured in the Dual Route Cascaded model of reading aloud (DRC) (Coltheart et al., 2001), see Figure 1.1. A non-lexical route stores GPCs, essential for novel word reading, and a lexical route stores whole-word orthographic forms and their pronunciations, essential for irregular word reading. The lexical route also enables us to recognise whether the visual form of a word is known or novel. Support for the DRC comes from research demonstrating that

adults read regular words faster than irregular words (Coltheart, 1978; Parkin, 1982; Waters & Seidenberg, 1985). The latency cost for irregular words occurs because the non-lexical route generates an incorrect regularised pronunciation which must be blocked by the lexical route. In contrast, for regular words there is no competition between the pronunciations generated by the two routes. In addition, the frequency of a word in written literature influences reading speed for irregular but not regular words (Paap & Noel, 1991; Seidenberg, Waters, Barnes, & Tanenhaus, 1984). This is because lexical representations of high frequency words are robust meaning that the lexical route blocks the activation of the non-lexical route quickly, reducing competition.

Further support for the DRC is provided by evidence from cognitive neuropsychology. Acquired brain damage can lead to phonological dyslexia, a specific deficit in nonword reading that is indicative of damage to the non-lexical route, but also the opposite, surface dyslexia, a specific problem with irregular word reading, suggesting damage to the lexical route (Marshall & Newcombe, 1973; Patterson, 1981). Castles and Coltheart (1993) argued that these patterns can also be observed in developmental dyslexia, with some children having greater problems with nonword reading than irregular word reading, and some having the reverse pattern. The data presented thus far support the DRC's suggestion that the mechanisms which represent spelling-sound rules (required for generalization) and item-specific knowledge (required to read irregular words and recognise the orthographic forms of known words) are separable.

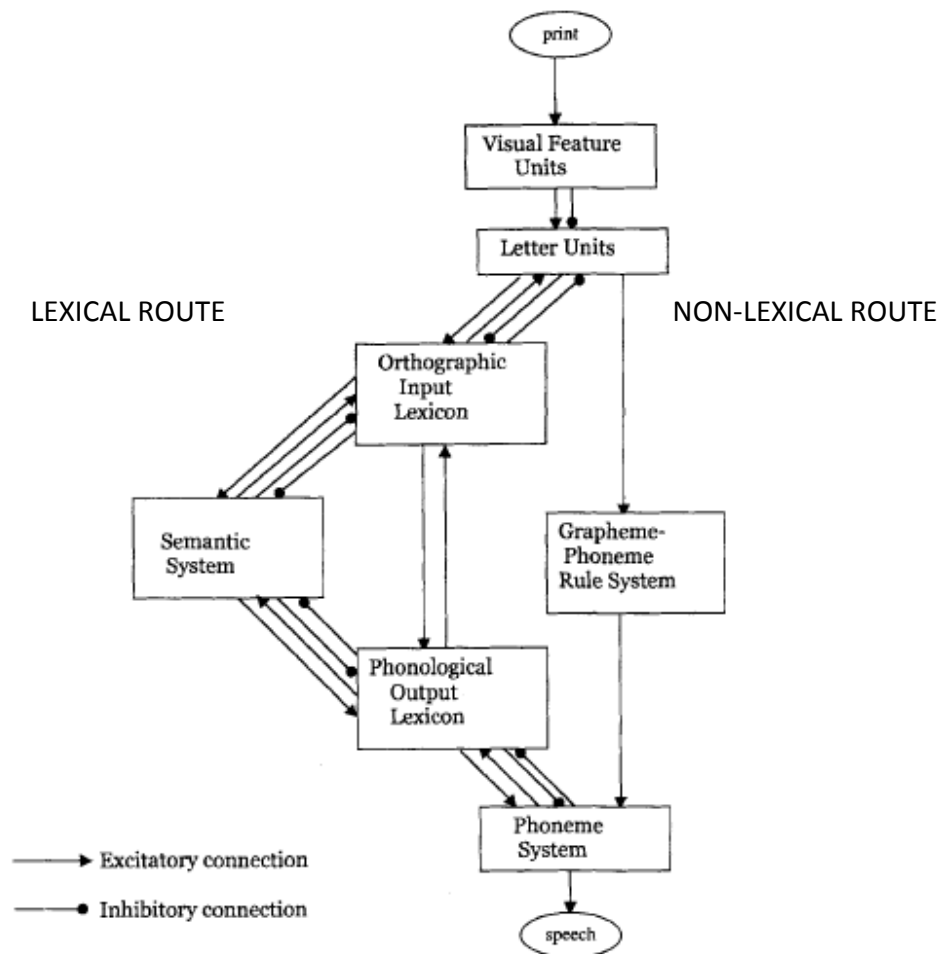


Figure 1.1. The Dual Route Cascaded model of reading aloud

The GPCs that characterise the non-lexical route of the DRC are not context-sensitive, i.e., they do not take into account the surrounding letters in a word. For example, GPCs do not capture the fact that, although the most common pronunciation of OO is /u/ as in SPOOK, OO is pronounced /u/ in the overwhelming majority of words that end in K, e.g., BOOK, TOOK, LOOK, COOK, ROOK, HOOK. Instead such words are classed as irregular and their pronunciations are represented in an item-specific way by the lexical route. However, evidence suggests that we are sensitive to context when reading aloud. Glushko (1979) found that *consistent* words, in which the orthographic

body (vowel plus following consonants) has only one possible pronunciation in English (e.g., RINK - SINK, LINK, PINK), were read faster than *inconsistent* words, which contain orthographic bodies with more than one possible pronunciation (e.g., FOOD – GOOD, HOOD, MOOD). This was the case even when all graphemes in the words were pronounced according to GPCs. Glushko's experiments suggest that our reading system represents more spelling-sound information than is captured by GPCs.

Jared, McRae, and Seidenberg (1990) refined Glushko's (1979) research and developed a graded measure of consistency. According to Glushko, BOOK is inconsistent because SPOOK is pronounced differently. However, as stated earlier, the majority of words ending in OOK are pronounced in the same way as BOOK, thus BOOK has a large number of friends (words with the same orthographic body pronounced in the same way) and only a single enemy (words with the same orthographic body pronounced differently). Jared et al. therefore classed BOOK as a highly consistent word and SPOOK as a highly inconsistent word. Some words also lie in between these extremes, for example, BLOWN has some friends, MOWN, FLOWN, GROWN, SOWN, and some enemies, CROWN, GOWN, TOWN, DOWN.

Jared et al. (1990) found that the magnitude of the effect of consistency on reading aloud was best captured by a proportional measure reflecting the number of friends vs. enemies a word possessed. Subsequent psycholinguistic research supports this position and has demonstrated that Jared et al.'s graded consistency measure captures more of the variance in reading aloud than regularity (Cortese & Simpson, 2000; Jared, 2002). In particular, a large scale regression study demonstrated that the

consistency of the orthographic body accounts for the greatest proportion of variance in the naming latency of English monosyllables (Treiman, Mullennix, Bijeljacbabic, & Richmond-Welty, 1995). These findings demonstrate that we are sensitive to conditional and probabilistic relationships between spelling and sound. This questions the idea that our reading system represents explicit pronunciation rules and divides words into those that do and do not obey these rules.

Consistency, like regularity, interacts with word frequency (how often a word is experienced in written literature). Jared (2002) found that consistency had no influence on error rates when reading aloud high frequency words, but, when reading low frequency words, error rates were reduced when they were highly consistent relative to when they were highly inconsistent. The reverse interaction also operated: frequency had no influence on the naming of highly consistent words, but for highly inconsistent words, error rates were lower when they were frequent than when they were infrequent. The interaction between frequency and consistency was also evident in naming latencies, although the effect was less reliable. These findings demonstrate that our reading system is sensitive to both the spelling-sound similarity between words, and the number of times a particular word has been experienced.

The strength of the influence of consistency and frequency on word processing is further illustrated by studies using lexical decision. Lexical decision involves making YES/NO responses to indicate whether presented stimuli are words (targets) or nonwords (distractors). Typically, accuracy is extremely high and target reaction time (RT) is the dependent variable of most interest. Frequency effects are robust in lexical

decision (Balota & Chumbley, 1984; Balota, Cortese, Sergent-Marshall, Spieler, & Yap, 2004; Cortese & Khanna, 2007) because the more often a word is experienced, the faster it can be recognised.

There has been some debate about the reliability of consistency effects in lexical decision. Several authors have reported that consistency does not affect lexical decision (Cortese, 1998; Jared, 1997; Jared et al., 1990), arguing that this is because spelling-sound mappings should not influence performance in a task in which no overt pronunciation is required. However, Stone, Vanhoy, and VanOrden (1997) demonstrated that consistency effects do emerge in lexical decision if stimuli are controlled for feedback consistency. A word is said to be feedback inconsistent if its orthographic body could have been spelled in a different way (e.g., MEAD could be spelled MEED or MEDE). In contrast, in the preceding discussion, inconsistent words have referred to cases in which the orthographic body has multiple possible pronunciations (e.g., MEAD could also be pronounced to rhyme with HEAD). Such words are more correctly termed feedforward inconsistent. Another way to think about these variables is that the mapping between spelling and sound is one-to-many in feedforward inconsistent words, and many-to-one in feedback inconsistent words. Stone et al. showed that when stimuli were feedback consistent (i.e., they contained orthographic bodies with only one possible spelling in English, e.g., POUCH or POUNCE), feedforward consistent words (e.g., POUNCE) were responded to faster and more accurately than feedforward inconsistent words (e.g., POUCH which could be pronounced to rhyme with TOUCH). Building on Stone et al.'s work, Lacruz and Folk (2004) reported that if feedback

consistency was controlled, feedforward consistency effects interacted with frequency in lexical decision. Overall, observations of feedforward consistency effects in lexical decision demonstrate that spelling-sound typicality can influence performance even in a task which does not demand the computation of phonology from orthography. Note that throughout this thesis, the term *consistency* will refer to the feedforward consistency of the orthographic body, unless otherwise stated.

So far it has been shown that our reading system stores knowledge about the spelling-sound typicality of words, that this influences reading and recognition of known words, and that it is difficult to classify this information as either obeying or not obeying explicit rules. The next consideration is whether this knowledge is also used productively. Several studies have demonstrated that spelling-sound consistency has an impact on pseudoword reading. For example, Glushko (1979) found that despite the fact that the novel words HEAF and HEAN can both be pronounced using GPCs, HEAF took longer to read than HEAN, because -EAF is an inconsistent orthographic body (LEAF, DEAF), whereas -EAN is consistent (LEAN, DEAN, MEAN, etc.). Treiman, Kessler, and Bick (2003) used a slightly different method and examined the pronunciations adults gave to novel words containing vowel graphemes that are known to vary in pronunciation. They found that adults applied context-sensitive consistency information when reading such novel words. For example, VOOK was more often read to rhyme with BOOK, TOOK, and LOOK, than SPOOK, whereas VOOT was more often read to rhyme with BOOT, HOOT, and LOOT, than FOOT. This shows that we do not simply use a rule when pronouncing graphemes such as OO, as suggested by the DRC. Instead we seem to apply our

knowledge of how graphemes are pronounced in the context of other letters, across the words we have experienced. This provides some evidence that the item-specific information that we store about word pronunciations is also used in generalization.

The graded influence of consistency on word and nonword reading and visual word recognition suggests that the DRC may not adequately capture the way in which we represent spelling-sound mappings. Alternatives to the DRC are Parallel Distributed Processing models such as the triangle model of reading aloud (Harm & Seidenberg, 2004; Plaut et al., 1996), see Figure 1.2. The triangle model comprises sets of units coding for phonological (sounds), orthographic (letters) and semantic (meaning) information. The model learns to read by being presented with the orthographic form of a word, producing a pronunciation attempt and receiving the correct pronunciation as feedback. This feedback then modifies the strength of the connections between units, thereby increasing the probability that future pronunciation attempts will be correct.

Importantly, the triangle model has no inbuilt GPCs and develops context-sensitivity, pronouncing nonwords such as VOOK to rhyme with BOOK, TOOK, and LOOK, rather than SPOOK. Furthermore, the model captures the frequency by consistency interaction seen in adult readers (Jared, 2002; Paap & Noel, 1991; Seidenberg et al., 1984). However, this is not a result of competition between dual mechanisms, as in the DRC. Instead, it is a consequence of the fact that inconsistent words suffer inhibition from other words with similar spellings but different pronunciations. If experienced frequently, the spelling-sound mappings for inconsistent words become strong enough to counteract this inhibition. In contrast, consistent words suffer from less inhibition,

because most words are pronounced in a similar way, and are therefore less influenced by frequency.

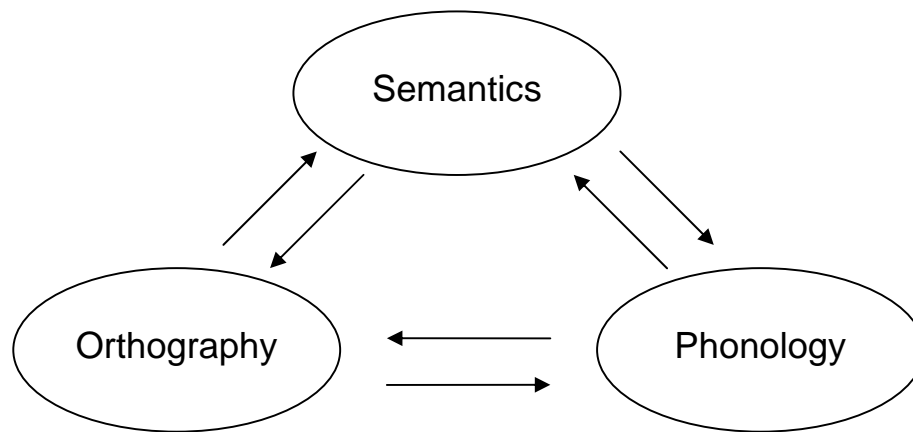


Figure 1.2. The triangle model of reading aloud

A point that should be made here is that, as discussed by Plaut et al. (1996), neither the consistency of the orthographic body nor regularity (whether or not a word is pronounced according to GPCs) adequately capture the “quasi-regularity” of English orthography. In one of the few studies to consider variability in spelling-sound mappings at multiple levels, Treiman et al. (1995) found that naming latencies were also influenced by the consistency of the onset plus vowel. For example, WAND suffered less from the inconsistency of its orthographic body (LAND, SAND, HAND), because many words beginning with WA are pronounced in the same way (WANT, WAD, WATCH, WAS). However, researchers studying the behaviour of models and human participants typically quantify spelling-sound typicality according to regularity or orthographic body consistency. When discussing the existing literature throughout this thesis, the use of

the terms regularity and consistency will reflect those adopted in the particular experiment being described.

The triangle model does not dedicate different units to the representation of different word types (i.e., words vs. nonwords, or regular words vs. irregular words). Instead, effects of lexicality, frequency, and spelling-sound context-sensitivity emerge as a consequence of statistical learning from exposure to the language. This model suggests that knowledge of context-sensitive spelling-sound mappings develops through exposure to whole words and their pronunciations. This is in contrast to the DRC which is a static model and avoids issues of learning on the grounds that "... unless the learning procedure itself is known to be psychologically real, it may not be able to learn what people learn..." (Coltheart et al., 2001, p. 216). It is true that the learning mechanisms employed in models may not be the same as those used by the human reading system. However, examining learning is essential to developing a comprehensive understanding of the operation of this system. Models of reading aloud should be considered to be theories to be evaluated against empirical data, whether in terms of the developmental or the mature state. The failure of the DRC to consider development therefore reduces its utility as a theory because it generates no predictions about learning. In contrast, the triangle model generates the explicit predictions that context-sensitive knowledge of spelling-sound mappings, and sensitivity to the frequency and consistency of these mappings, should emerge through exposure to whole-word orthographic forms and their pronunciations. These predictions are evaluated in Experiments 2 and 3 of this thesis (reported in Chapters 3 and 4).

The influence of word meanings

One important factor that has not yet been discussed is the influence of word meanings (semantics) on reading aloud and visual word recognition. In the DRC model, known words (both regular and irregular) can be processed either by a semantic or a non-semantic lexical route. This means that word meanings have the opportunity to influence our ability to read and recognise all known words. As the semantic lexical route has not been implemented, however, it is difficult to make explicit predictions as to how a semantic contribution might operate.

In contrast, the triangle model (Plaut et al., 1996) makes specific predictions about the role of semantics in reading aloud. In this model, the sets of units that encode phonological and orthographic information are supported by a bank of units intended to represent semantics. This semantic support provides additional input to the phonological units which pushes them towards their correct activations. Simulations by Plaut et al. (1996) demonstrated that the pathway from orthography to phonology via semantics was utilised more when reading low frequency inconsistent words. Such words have atypical and uncommon spelling-sound mappings and therefore pose considerable difficulty for the direct orthography-phonology route. A boost from semantics increases the system's efficiency when reading these words.

As Plaut et al. (1996) acknowledge, the way in which semantics was implemented is impoverished and does not capture how variations in word meaning might influence learning. However, it did simulate the idea that learning about orthography-phonology mappings does not occur in isolation and that semantic support

should aid this process. Furthermore, a later version of the triangle model (Harm & Seidenberg, 2004) did implement distributed semantic representations and produced many of the same results reported by Plaut et al., such as the interaction between frequency and consistency and the greater reliance of inconsistent words on the semantic route. However, as the focus of these later simulations was on reading comprehension, Plaut et al.'s work provides a more detailed consideration of the role of semantics in reading aloud. Throughout this thesis I will use the term *the triangle model* to refer to the general framework described in the preceding paragraph and outlined in both the 1996 and 2004 papers, with specific citations when appropriate.

Research findings provide some support for the triangle model's proposed role for semantics in reading aloud. Graham, Hodges, and Patterson (1994) and Patterson and Hodges (1992) observed that patients with semantic dementia exhibited specific deficits in reading words with atypical spelling-sound correspondences, despite performing normally on words with typical spelling-sound correspondences, and nonwords. This supports the idea that semantics aids irregular/inconsistent word reading. More recent neuropsychological data has demonstrated that difficulties in inconsistent word reading in semantic dementia are most pronounced for low frequency words (McKay, Castles, Davis, & Savage, 2007; Patterson et al., 2006; Woollams, Lambon-Ralph, Plaut, & Patterson, 2007). However, there have also been observations of intact inconsistent word reading in patients with semantic dementia (Blazely, Coltheart, & Casey, 2005; Schwartz, Marin, & Saffran, 1979). Furthermore, it is difficult to draw firm conclusions from neuropsychological evidence because it is always possible

that damage extends to other language regions, such as those that represent orthography-phonology mappings.

Corroborating evidence for a role for semantics in reading aloud is provided by research on typical adults. Strain, Patterson, and Seidenberg (1995) constructed lists of words that varied in imageability; "... the extent to which the representation of a word's meaning has sensorimotor properties..." (p. 1140). Imageability did not influence naming speed for high frequency words, or words with regular and consistent spelling-sound mappings; but, when reading low frequency exception words, adults were reliably slower and more error prone when they were low in imageability, relative to when they were high in imageability. This finding was replicated by Frost et al. (2005) who also demonstrated that brain activity associated with the difficulty of reading inconsistent words was modulated by imageability: high imageability inconsistent words produced less activity than low imageability inconsistent words. This was interpreted as an indication that semantic support reduced the difficulty of inconsistent word reading. In a regression analysis, Strain, Patterson, and Seidenberg (2002) obtained similar behavioural results, finding that imageability predicted naming speed for low frequency inconsistent words. These findings suggest that word meanings play a direct role in the computation of phonology from orthography and that this is most important when orthography-phonology mappings are uncommon and somewhat atypical.

Visual word recognition research also supports the view that semantic variables impact on orthographic processing. Whaley (1978) conducted a factor analysis of the variables influencing lexical decision and then used these factors in a multiple regression

analysis. A factor which included concreteness (the extent to which a word refers to objects, persons, or materials, rather than concepts which could not be experienced by the senses), imageability, and meaningfulness (how many associated words can be generated in a short time period) accounted for significant variance in decision time, over and above word frequency and length, confirming that semantic variables can play a role in visual word recognition. Chumbley and Balota (1984) found that a different measure of semantics, the speed with which participants could generate an associated word for each target, again predicted lexical decision time, over and above word frequency and length. More recently, a large scale regression study replicated the effects of imageability and meaningfulness on lexical decision and also found semantic connectivity (the number of other words with which a word is associated) to be predictive of response time (Balota et al., 2004). Note that none of these studies have reported an interaction between semantic effects and spelling-sound consistency or word frequency in lexical decision.

Lexical decision has received relatively little attention in connectionist modelling research using the triangle model framework (Harm & Seidenberg, 1999;2004; Plaut et al., 1996). Plaut (1997) proposed that the triangle model could make word-nonword judgements on the basis of the orthographic, phonological, or semantic familiarity of a stimulus, and indeed any combination of these sources of information. He then went on to demonstrate that a version of the triangle model could make highly accurate lexical decisions on the basis of semantic familiarity alone. This supports the idea that a system which does not possess whole-word representations can nevertheless recognise familiar

forms. As semantic and orthographic representations are reciprocally connected, the triangle model framework clearly suggests that word meanings may impact on visual word recognition. Unfortunately, however, modelling investigations have not directly addressed interactions between the effects of frequency, consistency, and semantics on lexical decision.

To summarise, behavioural and neuropsychological research provides some support for the role of semantics in orthographic processing embodied in the triangle model. The semantic properties of words seem to influence our ability to read them aloud, particularly when words are low in frequency and have somewhat atypical spelling-sound mappings. Word meanings also influence our ability to recognise them as familiar forms. This proposed role for semantics in orthographic processing is evaluated in Experiments 1, 4, and 5 of this thesis (reported in Chapters 2, 5, and 6).

Problems with existing research

Existing research suggests that when reading and recognising words we are sensitive to their frequency, spelling-sound consistency, and meaning. This supports the predictions of the triangle model (Harm & Seidenberg, 2004; Plaut et al., 1996) that our knowledge of spelling-sound mappings is sensitive to frequency and similarity, and that semantic information plays a direct role in orthographic processing. However, the research described so far has examined how we process words in natural languages. Such research suffers from several methodological problems which will now be explored.

One problem is how to quantify variables of interest. For example, frequency estimates are typically objective measures based on large corpora of written words, such as the Brown Corpus (Kucera & Francis, 1967), and the Center for Lexical Information (CELEX) database (Baayen, Piepenbrock, & Gulikers, 1995). However, these may not sufficiently capture the effect of word frequency on orthographic processing. Balota et al. (2004) found that subjective frequency accounted for significant variance in naming and lexical decision speed, over and above standard objective frequency measures. In a naming task, Zevin and Seidenberg (2004) found this also to be the case for cumulative frequency. Studies that do not control for such variables may underestimate the influence of frequency on orthographic processing, and as a result, risk overestimating the influence of other variables.

Similar issues arise when trying to operationalize semantics. In their experiments investigating the influence of semantics on word naming, Strain et al. (1995; 2002) manipulated imageability. However, as described earlier, research suggests that concreteness, meaningfulness, and the number of semantic associates also predict naming speed (Balota et al., 2004; Chumbley & Balota, 1984; Whaley, 1978). Furthermore, the factors that influence performance may differ between tasks. Balota et al. found that naming speed was influenced by imageability, but that lexical decision was also affected by meaningfulness and number of associates. In addition, the direction of semantic effects is not always easy to predict. Hino, Lupker, and Pexman (2002) demonstrated that words with ambiguous meanings elicit faster responses in lexical decision, but that words with many synonyms produce slower responses. These

findings illustrate that words are likely to vary along many other semantic dimensions than the one or two that are being manipulated. As with word frequency estimates, this means that experiments may fail to capture the variance in performance associated with word meaning.

Embedded in this operationalization problem is the fact that measures of variables such as frequency, consistency, and semantics are taken from lexical databases. For example, words are classified as consistent or inconsistent on the basis of their similarity to the spelling-sound mappings of other words in a database. However, participants in an experiment may not have experienced all the words in the database and they may therefore differ in the extent to which they consider a word to be consistent or inconsistent. This is most clearly illustrated by considering a hypothetical case of a young reader who has only seen two words ending in *-AVE*; *GAVE* and *HAVE*. To this child, both these words are equally consistent/inconsistent, whereas to an adult, *HAVE* is highly inconsistent. Thus, lexical database estimates are at best only a proxy for individual experience.

A third problem is that many experiments have used factorial designs that artificially divide continuous variables into discrete categories. Consistency, for example, should capture whether a word is pronounced in the same way as similarly spelled words and this should not just be with reference to the orthographic body. However, as Plaut et al. (1996, p. 59) acknowledge, "... if experimenters had to consider consistency across orthographic neighborhoods at all possible levels... their selection of stimulus words would be an even more agonizing process than it already is...". Although

understandable for practical reasons, artificial divisions can be problematic on several counts. First, it can mean that researchers construct somewhat different categories, making cross-study comparisons difficult. For example, in their experiments critiquing Strain et al. (1995), Monaghan and Ellis (2002) classified BEAD as inconsistent because of the 15 words ending in -EAD, five rhyme with BEAD and ten rhyme with HEAD. However, Strain et al. (2002) point out that this is a relatively small difference and argue that the inclusion of such stimuli may have reduced the influence of consistency (and its interaction with other variables) in Monaghan and Ellis's experiments.

Factorial designs may also overestimate differences between groups because category variables may be correlated (MacCallum, Zhang, Preacher, & Rucker, 2002). In Balota et al.'s (2004) large scale regression analysis, imageability correlated with both frequency and consistency and the interaction between the effects of imageability and consistency on naming latencies was not reliable. It might therefore be the case that in using a factorial design which failed to consider the correlation between imageability and consistency, Strain et al. (1995) inflated the strength of the interaction between these two variables. A final issue with such designs is that they encourage restricted and potentially unrepresentative item lists. For example, Ellis and Monaghan (2002) found that when a single item (COUTH) was removed from a word list used by Strain et al. (2002), the key theoretically important interaction between imageability and consistency disappeared. Overall these findings suggest that reliance on factorial designs may have caused studies to yield unreliable effects of frequency, consistency, and semantics on reading aloud.

This leads on to a further issue: how to select appropriate stimuli. Forster (2000, p. 1109) points out that although "... few researchers would deny the importance of random selection..." in fact "...each experimenter hand picks the materials, often with an elaborate but unarticulated set of criteria about what kinds of items are appropriate for an experiment..." This is because it is often deemed necessary to exclude certain types of words, such as proper names, inflected forms, number and function words, slang, words connected with current topical events or with sexual connotations, or words with common prefixes. Although excluding unsuitable items may not be problematic in itself, it introduces the opportunity for word lists to be selected to minimise or maximise the likelihood of obtaining significant effects of particular variables of interest. Forster demonstrated that word recognition researchers were able to predict which of two words would produce the fastest response, even when these words were equated for frequency. It is clear that there is potential for experimenter bias in stimulus selection due to pre-existing expectations of how variables should influence behaviour.

Another major concern is that there are many non-controlled factors which may influence performance. As described earlier, Lacruz and Folk (2004) found that feedforward consistency interacted with frequency in lexical decision only when feedback consistency was controlled. This provides an example of how failing to control for/measure one variable can mask the effects of others. The problem of non-controlled factors has been particularly controversial in research on semantic effects on reading aloud. Gilhooly and Watson (1981) demonstrated that reading aloud is influenced by the age at which words are acquired. Monaghan and Ellis (2002) reanalysed Strain et al.'s

(1995) data and found that when AoA was controlled, imageability no longer affected word naming. Similar findings were reported by Cortese and Khanna (2007) in a reanalysis of Balota et al.'s (2004) large scale regression study. Age-of-acquisition (AoA) is correlated with imageability because we tend to learn high imageability words, such as SHOE or BRIDGE, earlier than low imageability words, such as RATE or LURE. These findings therefore suggest that imageability effects may in fact be driven by a reading aloud advantage for words learned early in life.

The preceding discussion has highlighted methodological problems arising from the correlations among psycholinguistic variables. However, the correlation between AoA and imageability in fact turns out to have important theoretical implications. Several authors have argued that the influence of AoA on orthographic processing arises because early learned words have stronger phonological representations (Brown & Watson, 1987; Morrison & Ellis, 1995; Morrison, Ellis, & Quinlan, 1992), although others suggest that both semantic and phonological familiarity may contribute to AoA effects (Bonin, Barry, Meot, & Chalard, 2004; Cortese & Khanna, 2007). If AoA accounts for imageability effects, and the influence of AoA is in part driven by differences in phonological representations, imageability effects may to some extent reflect the influence of a phonological variable. This highlights a specific problem for research investigating the role of semantics in orthographic processing: at least some of the reported effects associated with familiarity with word meanings may in fact arise from familiarity with word sounds.

A rather different issue with existing research is that it largely fails to address the issue of how frequency, consistency, and semantics influence the process of *learning* to read. Developmental research has shown that children use grapheme-phoneme correspondence rules and are sensitive to consistency when reading aloud words (Weekes, Castles, & Davies, 2006) and pseudowords (Stuart, Masterson, Dixon, & Quinlan, 1999; Treiman, Goswami, & Bruck, 1990). However, it is difficult to determine the extent to which this information is explicitly taught vs. implicitly extracted through exposure. Thus, these studies do not tell us *how* children learn spelling-sound information. Regarding the role of semantics, Ricketts, Nation, and Bishop (2007) demonstrated that in 9-year-old children, a standardized measure of expressive vocabulary was predictive of exception word reading. This supports the idea that children use semantic knowledge when reading words with atypical spelling-sound mappings. However, as in the adult literature, it is possible that this relationship was driven not by semantic familiarity but by phonological familiarity. This is because the more familiar a child is with the meaning of a word, the more familiar they are likely to be with how it sounds. For example, if a child remembers to pronounce HAVE /hæv/ rather than /herv/, is this because they know the meaning of the word, or because they are familiar with the phonological form of the former (but not the latter). The developmental literature will be discussed more fully in Chapters 2 and 3, but the above examples illustrate how the same problems plague both developmental and adult psycholinguistic approaches. They also demonstrate that we know relatively little about the process of orthographic learning.

To summarise, models of reading aloud make assumptions about how we learn and represent spelling-sound mappings. They also make predictions about how frequency, spelling-sound typicality, and semantics should influence learning and subsequent processing. These predictions are constrained by neuropsychological and behavioural data. However, the research that has generated this data suffers from several methodological problems, such as difficulties in measuring and controlling for lexical variables, artificial categorisation of graded variables, the use of potentially idiosyncratic item lists, and the difficulty of separating effects of phonological and semantic familiarity. Furthermore, existing research has failed to address how we learn about spelling-sound mappings and how this process is influenced by word meanings.

New approaches

One way to overcome the difficulties outlined in the preceding section is to expose participants to a novel language environment. This minimises the difficulty of measuring and categorising existing lexical variables and reduces concerns about the influence of non-controlled factors. Artificial language learning paradigms have been used successfully in spoken language processing research, some examples of which will be described in the next section. These examples are not exhaustive, but illustrate the utility of using novel language environments. The focus then turns to some recent experiments in which adults learn to read novel words written in English orthography. Following this is a discussion of how to assess the extent to which newly learned items

are processed similarly to words in natural languages. Finally, consideration is given to experiments which have exposed learners to a novel orthography.

Learning new spoken words

Magnuson, Tanenhaus, Aslin, and Dahan (2003) used an artificial language learning paradigm to examine the development of competition effects during spoken word recognition. Models of spoken word recognition suggest that cohort competitors (words differing only in the final phoneme) produce greater interference than rhyme competitors (words differing only in the onset) because word recognition is to some extent a serial process. However, some models predict that rhyme competition should only occur when there is minimal dissimilarity between onsets (e.g., pet vs. bet) (Marslen-Wilson, 1987; Norris, 1994), whereas others suggest that rhyme competition will always occur, albeit to a minimal extent (McClelland & Elman, 1986). Furthermore, as children may have more holistic lexical representations (e.g, syllable-based rather than segment-based) (Charles-Luce & Luce, 1990), it might be predicted that rhyme competitors should produce stronger interference in children than in adults. Magnuson et al. argued that natural language investigations of these issues are compromised by their use of lexical database estimates of phonological overlap because "... forms listed in a corpus will differ substantially from the context-dependent surface realization of spoken words in fluent speech..." (p. 203).

To overcome these stimulus control problems, Magnuson et al. (2003) exposed adults to new spoken word forms that varied in phonological overlap. At the end of training, cohort competitors produced stronger interference than rhyme competitors,

but rhymes produced weak interference effects throughout, even when onsets were phonologically dissimilar. This supports McClelland and Elman's (1986) model of spoken word recognition. Furthermore, although rhyme competitors did produce stronger interference than cohort competitors early in learning, there was also evidence of incremental processing from the beginning of training. This suggests that differences in child and adult processing may not be due to qualitative differences in the organisation of the lexicon (i.e., holistic vs. segmental representations), but to quantitative differences in the strength of lexical representations. Magnuson et al.'s artificial language paradigm provided a high degree of experimental control over the phonological properties of the novel words and enabled them to evaluate processing during early lexical learning.

A similar method was used by Creel, Aslin, and Tanenhaus (2006) to examine the relative importance of consonants and vowels for word identity. Previous research had indicated a consonant superiority effect (Cutler, Sebastian-Galles, Soler-Vilageliu, & Van Ooijen, 2000; Marks, Moates, Bond, & Stockmal, 2002; Van Ooijen, 1996), but such research is confounded by the fact in the languages studied (English, Spanish, and Dutch) most words begin with consonants. Creel et al. found that adults did indeed value consonant identity over vowel identity when distinguishing between CVCV words, but that vowel identity was more important in VCVC words. They concluded that initial segments, rather than consonants are of primary importance for word identity.

An artificial language learning paradigm has also been used to evaluate a problem in the domain of grammatical learning. Wonnacott, Newport, and Tanenhaus

(2008) exposed adults to a new language in which verbs probabilistically occurred in one of two constructions. Post-tests demonstrated that adults had learned and abstracted the statistical regularities governing verb use that were implicit in the language, and could use them productively and in on-line comprehension. Previous work had been hampered by the fact that semantic and syntactic properties of verbs are highly correlated and thus had been unable to establish whether learners were sensitive to one or the other or both. This work established definitively that learners are able to extract and use distributional syntactic information about verb usage.

These examples illustrate how artificial language learning paradigms can be used to complement and extend natural language research. The methods used provide greater control over exposure to the statistical properties of the language environment, enabling researchers to tease apart the effects of variables that are often correlated in natural languages. They also provide a window on early language learning processes, something that is difficult to achieve even in developmental studies.

Learning to read new words

Artificial language learning techniques have been adapted to examine visual word processing by exposing learners to novel words written in English orthography. Bowers, Davis, and Hanley (2005) used such a method to investigate the validity of competition based models of visual word recognition (e.g., McClelland & Rumelhart, 1981). Such models suggest that words with similar orthographic representations will compete with each other for recognition. The prediction then follows that words with one or more neighbours (words that differ in a single letter), will take longer to

recognise than words with no neighbours, so-called hermits. Existing research has provided mixed support for this prediction (for a review see Andrews, 1997) and Bowers et al. suggested that this may in part be due to failures to match stimuli on dimensions such as age-of-acquisition and imageability. To overcome this, they taught adults new novel forms, such as BANARA, that were related in form to hermit words (e.g., BANANA). Participants made semantic category judgements in response to the orthographic forms of hermit words, both before and after they had learned the novel words. Bowers et al. found that judgements about hermit words were slower after novel word training. Because target words were compared with themselves (before and after novel word learning), the stimulus matching problem was avoided. The experiment provided clear support for the process of lexical competition embodied in many models of visual word recognition.

Recently, two novel word learning experiments have investigated the role of semantics in orthographic processing. Trudeau (2006) trained adults to read monosyllabic novel words which were assigned pronunciations that were either consistent or inconsistent with the pronunciation of similar words in English. For example, MAVE pronounced to rhyme with GAVE (consistent), or GLORM pronounced to rhyme with WORM (inconsistent). Half the consistent and half the inconsistent novel words were associated with a low imageability novel definition, for example, “the stress of having no money” and half of each set were associated with a high imageability novel definition, for example, “sausage made with fish meat”. Post-training, inconsistent

words were named faster when they referred to high imageability definitions, whereas consistent words were unaffected by imageability.

McKay, Davis, Savage, and Castles (2008, Experiment 2) used a similar method to investigate the contributions of phonological and semantic familiarity to orthographic processing. As discussed earlier, these two variables are confounded in natural language research because the more familiar we are with a word's meaning, the more familiar we are likely to be with its phonological form. Adults were familiarised to the sounds of 20 novel words (lexical phonology pre-exposure) and to definitions for half of these words (semantic pre-exposure), prior to learning to read them. As in Trudeau's (2006) experiment, half the novel words were trained with consistent pronunciations and half with inconsistent pronunciations. Semantic pre-exposure improved subsequent reading accuracy and speed relative to lexical phonology pre-exposure, but only when the novel words had inconsistent pronunciations. Taken together, these two studies support the specific role for semantics in reading aloud suggested by the triangle model. Importantly, whereas previous natural language research on imageability effects (Strain et al., 1995;2002) was confounded by the correlation between AoA and imageability (Ellis & Monaghan, 2002; Monaghan & Ellis, 2002), the method used in these studies ensured that AoA was held constant. Furthermore, McKay et al.'s experiment provides the first direct evidence that semantics can influence word reading processes over and above any possible effects of phonological familiarity.

Although the experiments reported by Trudeau (2006) and McKay et al. (2008) seem to provide convincing evidence for a role for semantics in reading aloud, there are

several issues that should be mentioned. First, both studies used a within subject design and this may have caused artefactual differences between conditions. In Trudeau's experiment, each participant learned high imageability definitions for half the words and low imageability definitions for the other half of the words. As high imageability definitions are likely to be more memorable, it may be that participants paid minimal attention to the low imageability items. It is difficult to assess how well high and low imageability definitions were learned because semantic learning was tested with a non-demanding two-alternative forced choice task. Similarly, in McKay et al.'s experiment, participants learned definitions for half the items and simply listened to and repeated the phonological forms of the remaining items. Thus, it is possible that attention was focused on the difficult definition learning task and directed away from the listen-repeat task. This means that McKay et al. may have underestimated the role of phonological familiarity, and in turn overestimated the role of semantic familiarity, in reading aloud.

A second issue is that although McKay et al. (2008) demonstrated that semantic familiarity enhanced reading aloud over and above phonological familiarity, the role of phonological familiarity is in fact difficult to evaluate because a baseline no familiarity condition was not included. Thus, it may be that familiarity with word sounds does support reading aloud, albeit to a lesser extent than familiarity with word meanings. Finally, because novel words were written in English orthography, reading accuracy was high from the beginning of training in both Trudeau (2006) and McKay et al.'s experiment. Thus, neither shed light on whether/how knowledge of word meanings supports the early stages of learning to read. Overall, although these novel word

learning experiments have enabled factors such as frequency, imageability, and AoA to be tightly controlled, the role of semantics in orthographic learning warrants further investigation.

How do we treat newly learned words?

McKay et al. (2008) sought to verify that newly learned words were processed similarly to real words. They suggested that this would increase the relevance of their findings to understanding word reading processes in natural languages. First, they examined whether semantic pre-exposure influenced decision speed in a task requiring participants to judge whether a presented item was a trained item or an untrained but similar item. This old-new decision task was based on typical lexical decision; recall that semantic variables affect response latencies in lexical decision (e.g., Balota et al., 2004; Cortese & Khanna, 2007). McKay et al. found that old-new decisions to trained words were faster following semantic pre-exposure than following lexical phonology pre-exposure. This suggests that trained items were processed similarly to real words. It also demonstrates that semantic familiarity influences visual word recognition as well as reading aloud, consistent with previous research and the predictions of the triangle model.

The second measure that McKay et al. (2008) used to index lexicalisation was whether masked identity priming influenced the speed with which participants could name trained items. In natural languages, naming and lexical decision RTs are decreased by prior presentation of the same word, relative to a different word. Forster and Davis (1984) demonstrated that, in lexical decision, if the prime is of short duration and is

masked such that it is not consciously perceived, this *masked identity priming* effect is reliable for words but not nonwords. On this basis, McKay et al. argued that if masked identity priming effects were specific to trained (and not untrained) items in naming, this would provide further evidence that trained items were treated similarly to real words. McKay et al. found that trained novel word naming latencies were facilitated by masked identity priming relative to control priming. The size of the effect was equivalent to that seen for real words, and was not present for a further set of untrained novel words. McKay et al. took this as evidence that trained words were incorporated into the existing lexicon and argued that this increased the relevance of their findings for understanding orthographic processing in natural languages.

However, there is in fact much debate about the lexical nature of masked priming effects. Proponents of the lexical view (e.g., Forster, Mohan, & Hector, 2003; Monsell, 1991; Morton, 1969) argue that priming exerts its influence by activating existing lexical representations. This decreases RTs to word targets because activation levels are nearer the threshold required for a response to be made. Nonword distractors should not show this effect because they do not have pre-existing lexical representations. An alternative to this lexical view is that primes create an episodic memory trace that facilitates responses to the target because it requires similar (or identical) processing (Masson & Bodner, 2003). Thus, priming effects should be evident for words and nonwords because both are able to create a memory trace.

Forster (1998) argued that there must be a lexical component to masked priming in lexical decision because effects still occur when there is no visual similarity between

prime and target. For example, in bilingual participants, masked priming occurs between translation equivalents even when the two scripts are completely different (Gollan, Forster, & Frost, 1997). Similarly, Bowers, Vigliocco, and Haan (1998) demonstrated that lower case identity primes facilitate responses to uppercase word targets even when the two forms are visually distinct, for example, read-READ, head-HEAD. These findings support the idea that priming can have a lexical origin, but do not rule out an episodic contribution.

Forster (1998) reviewed studies of masked priming effects in lexical decision and concluded that priming effects are strong and reliable for words, with the mean facilitation in latency being 50-60ms, whereas nonword priming effects are either absent or much smaller, averaging around 8-10ms facilitation. This suggests a large lexical component for masked priming in lexical decision. However, Bodner and Masson (1997) argue that nonword priming effects in lexical decision are negated because identity priming increases nonword familiarity and therefore the likelihood of a false positive response. They demonstrated that in tasks where responding on the basis of familiarity is reduced, for example, when stimuli are presented in MiXeD cAsE lEtTeRs, nonword priming effects are robust. However, Forster argued that this method explicitly focuses participants on the letter level, artificially causing letter-level priming (Jacobs & Grainger, 1991) which otherwise would not have occurred, and which does not drive typical masked priming effects.

Further evidence for a lexical basis to masked priming is offered by Forster, Mohan, and Hector (2003). In this lexical decision experiment, nonwords were

orthographically illegal, e.g., BCRWE. This meant that lexicality judgements could be made by deciding whether a combination of letters was orthographically familiar, rather than by evaluating whether it constituted a word. This minimised the necessity of accessing lexical representations which should remove priming effects if they have a lexical basis. Forster et al. found no evidence for nonword priming in this experiment. This was also the case for low frequency words. However, priming effects were obtained when words were high frequency. Forster et al. suggested that high frequency words have such strong lexical representations that they are activated even in a task designed to minimise lexical access. Thus, this study provided no evidence for non-lexical priming and strong evidence for lexical priming.

Overall, the evidence suggests that priming effects are much greater for words than nonwords in lexical decision. This might be taken as support for McKay et al.'s (2008) argument that because priming effects were specific to trained item naming this constitutes evidence that newly learned items were lexicalised. Unfortunately, however, the lexical basis of masked priming effects in naming is not well supported. Naming has received little attention in the priming literature with researchers predominantly using lexical decision. Davis, Kim, and Forster (2008) did report masked priming effects that were specific to words in a naming task. However, this was in a special case in which targets and distractors were presented backwards and primes forward. In a more typical naming experiment with forward primes and targets/distractors, Masson and Isaak (1999) found that priming effects in naming tasks are in fact robust for both words and nonwords. Forster et al. (2003) and Bowers (2003) draw similar conclusions in reviews of

masked priming. Thus, McKay et al.'s finding that masked identity priming increased naming speed for trained but not untrained novel words, does not in fact provide strong support for the lexicalisation of trained items in their novel word learning study. However, in lexical decision, masked priming effects do appear to be largely restricted to words suggesting that this might be a more fruitful way to assess lexicalisation in novel word learning experiments.

Learning a novel orthography

The experiments of Trudeau (2006) and McKay et al. (2008) provide some support for the argument that semantics plays a role in reading aloud, particularly for words with inconsistent spelling-sound patterns. However, they have not addressed the role of semantics in the early stages of orthographic learning because participants were highly proficient at reading the novel words from the beginning of training. Furthermore, although research suggests that semantic factors interact with word frequency and spelling-sound consistency in their effects on reading aloud, research to-date has yet to establish how sensitivity to these lexical variables develops. Experiments using familiar alphabetic scripts cannot shed light on this process because individual differences in prior experience with the orthography are likely to impact on novel word learning. How we learn spelling-sound mappings and the influence of frequency, consistency, and semantics on this learning process could be examined if participants were exposed to a novel orthography, but, unfortunately, very few studies have used such techniques. Those that have will now be reviewed.

In the 60's and 70's a cluster of studies emerged in which children or adults were trained on the associations between novel symbols and sounds (for a review see Knafle & Legenza, 1978). Similar methods were also used by Byrne (1984) and Byrne and Carroll (1989) to examine whether learners could extract phonetic feature information embedded in novel symbols. More recently these techniques have been combined with neuroimaging to look at changes in brain activity for novel vs. trained symbols (Callan, Callan, & Masaki, 2005; Hashimoto & Sakai, 2004). However, in all these studies the focus was on paired associate learning of single symbol to sound mappings. None investigated the process of learning sub-word spelling-sound mappings through exposure to whole words, or how this is influenced by frequency, consistency, and semantic factors.

Two experiments that have used word-level forms presented in an artificial orthography are described by Bitan and Karni (2003; 2004). They asked whether adults could extract sub-word spelling-sound correspondences from novel words written in a novel orthography and concluded that they could not. Although participants were able to learn the training sets, they were unable to generalize their knowledge to novel forms written in the same orthography without explicit teaching in symbol-sound correspondences. These findings contrast with the assumption embodied in the triangle model that sensitivity to spelling-sound patterns can be acquired through exposure to whole-word orthographic forms and their pronunciations. Instead they suggest that such knowledge must be acquired through direct teaching.

However, there are many differences between Bitan and Karni's (2003; 2004) methodology and the typical process of learning to read an alphabetic script. First, Bitan and Karni taught adults only 6-12 words; this may have encouraged a whole-word rote learning strategy. This may have been further exacerbated by the fact that, in contrast to the cohesive symbols used in most alphabetic scripts, each phoneme in Bitan and Karni's experiment was represented by two or three separate symbols, e.g., /[^]* or */. It was therefore fairly difficult to extract symbol-sound knowledge. Finally, training required participants to make decisions as to whether their decoding attempts matched an English word translation, and to judge whether pairs of trained items were the same or different. This learning process is very different to the corrected pronunciation attempts that typically characterise how children learn to read words. It may therefore be that learners can extract sub-word orthography-phonology mappings from a novel orthography if given an environment more typical to that of learners of alphabetic scripts.

To summarise, exposing participants to a novel language environment provides a greater degree of experimental control over lexical statistics than is possible in natural language research. It also minimises the problems associated with measuring and controlling for lexical variables. Furthermore, artificial language learning paradigms enable direct investigation of the learning process, as exemplified by the spoken word learning research described earlier. However, to-date, research on orthographic processing has not exploited this to the full.

In this thesis, novel word and artificial orthography learning experiments are used to address several important remaining questions. First, can we learn sub-word spelling-sound patterns through exposure to whole-word forms? Second, is knowledge extracted about spelling-sound mappings context-sensitive, i.e., what the other letters in a word are? Third, are orthographic learning, subsequent processing, and generalization influenced by the frequency and predictability of spelling-sound mappings present in the language environment, as is the case in natural languages? Fourth, to what extent is the same knowledge used to support item-specific (i.e., reading and recognition of trained items) and generalization processes (reading untrained items)? Finally, what is the role of word meanings in orthographic learning?

CHAPTER 2: SEMANTIC INFLUENCES ON ORTHOGRAPHIC PROCESSING:

A NOVEL WORD LEARNING EXPERIMENT

The triangle model of reading aloud (Harm & Seidenberg, 2004; Plaut et al., 1996) suggests that semantics supports word reading, particularly for items with inconsistent spelling-sound mappings. This experiment explored this prediction by teaching 80 adults to read 12 pseudowords with consistent pronunciations and 12 with inconsistent pronunciations. Prior to this, participants learned either high or low imageability definitions of the items (*semantic pre-exposure*), the sounds of the items (*lexical phonology pre-exposure*), or received *no pre-exposure*. Semantic pre-exposure increased reading accuracy for inconsistent items relative to no pre-exposure. Lexical phonology pre-exposure was less beneficial. These findings offer some support for the triangle model. However, pre-exposure did not decrease reading or recognition latencies. Furthermore, there was no advantage for learning high imageability, relative to low imageability, definitions, suggesting that the benefit of semantic pre-exposure may not be semantic in origin. It is concluded that new methods are needed to overcome several problems associated with research using English orthography.

As described in Chapter 1, the triangle model of reading aloud (Harm & Seidenberg, 2004; Plaut et al., 1996) suggests that the process of computing phonology from orthography is aided by knowledge of word meanings. Plaut et al. and Harm and

Seidenberg demonstrated that semantic support is particularly important when reading low frequency inconsistent words. This is because such words are most difficult for the direct orthography-phonology route to learn. Plaut et al. also showed that disruptions to the semantic route of the triangle model resulted in inconsistent word reading deficits. The role for semantics in reading aloud embodied in the triangle model is supported by findings from neuropsychology. This work has demonstrated that patients with semantic dementia have word reading deficits that are most pronounced for low frequency inconsistent words, and that the degree of semantic impairment predicts the extent of irregular/inconsistent word reading problems (McKay et al., 2007; Patterson et al., 2006; Woollams et al., 2007).

Data from experiments with typical adults provide further evidence for a role for semantics in orthographic processing. Strain et al. (1995) demonstrated that when naming low frequency words that were irregular and inconsistent, adults were faster and more accurate when the words were high imageability, for example, DOUGH or SWAMP, than when they were low imageability, for example, DREAD or SCARCE. However, imageability had no impact on words with regular and consistent spelling-sound patterns, or high frequency words. Strain et al.'s findings were replicated by Frost et al. (2005) and Woollams (2005). This work suggests that a word's semantic properties impact on the ease with which we can read that word aloud, particularly if it is uncommon and has an atypical spelling-sound pattern.

Developmental studies provide corroborating data. Nation and Snowling (2004) measured 8-year-old children's expressive vocabulary skills, one index of individual

differences in word-level semantic knowledge, and found that they predicted their exception word reading at age 14. However, as regular/consistent word reading was not assessed, this does not constitute direct evidence for an interaction between spelling-sound consistency and semantics in young readers. Such an interaction was demonstrated by Ricketts et al. (2007) who found that expressive vocabulary predicted accuracy of exception word reading, but not regular word or nonword reading, in a sample of 9-year-old children. In addition, this study found that children with reading comprehension problems had deficits in semantic knowledge and were also poor at reading exception words, but not regular words or nonwords. This mirrors the pattern of performance seen in patients with semantic dementia. These findings suggest that semantics influences inconsistent word reading fairly early in development.

Lexical decision experiments have demonstrated that semantic variables also impact on our ability to recognise familiar words. Whaley (1978) reported that concreteness, meaningfulness, and imageability influenced word-nonword decision times. The effects of imageability and meaningfulness were replicated by Chumbley and Balota (1984) and Balota et al. (2004), both of which also found the number of other words with which a word is associated to be predictive of response time. Although there has been relatively little explicit research using the triangle model framework into the impact of semantics on lexical decision, these findings are easily accommodated within this framework given that semantic and orthographic representations are reciprocally connected. It is important to note that, in contrast to the work on reading aloud, neither

modelling nor empirical research has reported an interaction between the effects of semantic variables and frequency or consistency on lexical decision.

Problems with existing research

The data discussed thus far seem to support the role for semantics in orthographic processing suggested by the triangle model. However, there are several reasons why this conclusion is not unequivocal. The evidence from neuropsychology is contentious because semantic deficits are not always associated with reading difficulties. For example, Blazely et al. (2005) described EM, a patient who showed severe semantic deficits but had completely intact word reading, including low frequency words with irregular spelling-sound patterns. Similarly, Gerhand (2001) reported that one patient with semantic dementia could accurately read all of Strain et al.'s (1995) low frequency exception words. Woollams et al. (2007) attempted to accommodate such findings within the triangle model framework on the basis of two predictions made by modelling simulations. First, differences are likely to exist in the extent to which typical individuals (or models) rely on the semantic route. In extreme cases, where use of the semantic system in computing phonology from orthography is minimal, damage to the semantic system may leave word reading intact. Second, inconsistent word reading problems should emerge as the severity of semantic impairment increases. These predictions were largely supported in a longitudinal study of 51 cases of semantic dementia. However, despite these arguments, the relationship between semantic and word reading deficits remains contentious.

Several criticisms can also be levied at research on typical adult reading. One problem is the selection of appropriate stimuli and the generality of results. For example, Ellis and Monaghan (2002) demonstrated that when a single item, COUTH, was removed from Strain et al.'s (2002) word list, the effect of imageability on low frequency inconsistent items disappeared. COUTH was removed because only 5 of the 24 participants read the item correctly, and, for those 5 participants, naming latencies were more than 4 standard deviations longer than the mean latency across items. Thus, imageability effects may be restricted to a very small number of items. This questions the overall importance of semantics for reading aloud.

As discussed in Chapter 1, a further issue with the research on typical adult reading is that researchers often fail to control for important variables. In a reanalysis of Strain et al.'s (1995) data and in similar experiments, Monaghan and Ellis (2002) found that imageability had no influence on naming once age-of-acquisition (AoA) was controlled. In response to Monaghan and Ellis (2002), Strain et al. (2002) conducted a multiple regression analysis on a larger word set and demonstrated that imageability, but not AoA, predicted naming latencies for inconsistent words. However, as described above, these results may have been driven by the failure to exclude an extreme outlier. In addition, Strain et al. (2002) failed to highlight that when naming *errors* rather than latencies were considered, AoA was a significant predictor and imageability was not. Similar findings of AoA underlying imageability effects were reported by Cortese and Khanna (2007) in a reanalysis of Balota et al.'s (2004) naming latency data. Overall,

these data seem to suggest that imageability effects on reading aloud may in fact be driven by AoA.

This discussion of the influence of AoA on reading aloud points to a third problem with current research on the role of semantics in orthographic processing: the correlation between semantic and phonological familiarity. Whilst imageability is by definition a semantic variable, it has been argued that AoA may at least in part be a phonological variable. For example, Brown and Watson (1987) suggested that AoA influences the integrity of phonological representations. This was on the basis of research demonstrating that AoA effects are only reliable in tasks requiring overt spoken output, such as word naming (Gilhooly & Watson, 1981). This view was supported by Morrison et al. (1992) who reported that AoA affects object naming but not object recognition, and Morrison and Ellis (1995) who found that AoA influences immediate but not delayed word naming. Subsequent discussions have acknowledged that AoA may influence phonological, orthographic, and semantic aspects of lexical representations (Bonin et al., 2004; Cortese & Khanna, 2007; Morrison & Gibbons, 2006; Morrison, Hirsh, Chappell, & Ellis, 2002). However, if AoA even in part exerts its effects on reading aloud because of its influence on phonological representations, the evidence that AoA may underlie imageability effects (Ellis & Monaghan, 2002; Monaghan & Ellis, 2002) weakens the argument that semantics plays a role in orthographic processing.

The problem of disentangling the influences of semantic vs. phonological familiarity can be further illustrated using developmental data. Expressive vocabulary has been used as a measure of word-level semantics and has been found to predict

exception word reading (Nation & Snowling, 2004; Ricketts et al., 2007), supporting the idea that semantics influences reading aloud. However, the more familiar a child is with the meaning of a word, the more familiar they are likely to be with its phonological form and it is not clear which of these factors is responsible for the relationship between vocabulary and reading aloud. For example, when reading an inconsistent word, such as WATCH, it might be that knowing its meaning helps us to pronounce it differently from MATCH, LATCH, and CATCH. However, it might simply be that possessing a representation of the sound of the word /wɒtʃ/ but not /wætʃ/ helps us to remember how to read this word. Thus, evidence from studies of children's reading also leaves open the possibility that semantic effects on orthographic processing are in fact driven by phonological familiarity.

Novel word learning studies

One way to overcome the problems of selecting appropriate stimuli and the correlation between AoA and semantic variables is to use an artificial lexicon. If participants are taught new words, variables such as frequency, consistency, imageability, and AoA can be controlled or manipulated. Many researchers have used such techniques to investigate spoken word learning but few have focused on visual word learning. Two studies that have which were outlined in Chapter 1 will now be described in detail. Trudeau's (2006) work aimed to re-evaluate Strain et al.'s (1995; 2002) argument that imageability influences inconsistent word reading. Adults learned to read 60 monosyllabic novel words written in English orthography. Each word was

presented in conjunction with a novel definition; 30 of these were of low imageability and 30 were of high imageability. Half the low imageability and half the high imageability novel words were assigned a highly consistent pronunciation, which rhymed with most English words with the same orthographic body, for example, RONK pronounced to rhyme with HONK. The remaining items were assigned a highly inconsistent pronunciation, which rhymed with only one or two English words with the same orthographic body, for example, SASTE pronounced to rhyme with CASTE. Post-training, both consistency and imageability influenced naming accuracy but there was no interaction between the two effects. However, in the latency analysis, inconsistent words were named faster when they were associated with high imageability definitions than when associated with low imageability definitions. Consistent word naming latencies were unaffected by imageability. This study demonstrates that imageability effects can occur independently of AoA and provides some support for the idea that semantics plays a specific role in supporting inconsistent word reading.

A similar methodology was used by McKay et al. (2008) to examine whether semantic familiarity benefits orthographic processing over and above phonological familiarity. Adults were trained to read 20 novel words written in English orthography. As in Trudeau's (2006) experiment, half of these words were assigned consistent pronunciations, and half inconsistent pronunciations. However, note that McKay et al.'s consistency manipulation differed from that used by Trudeau in that consistent pronunciations were entirely consistent and rhymed with *all* words in English with the same orthographic body, (e.g., TREN pronounced to rhyme with GLEN) and inconsistent

pronunciations were entirely inconsistent and *never* occur in English for that word body (e.g., SHILL pronounced to rhyme with MILE). Prior to learning to read the novel words, participants were familiarised with the sounds of all the items (lexical phonology pre-exposure) and with a novel definition for half the items (semantic pre-exposure). Results indicated that reading accuracy during training, and post-training naming accuracy and latencies, were facilitated by semantic pre-exposure, but only when items were inconsistent. In addition, participants were faster at discriminating trained from untrained novel words when they had been associated with definitions, although note that this effect was not specific to inconsistent items. These findings again support the role for semantics in orthographic processing suggested by the triangle model. In addition, they provide the first demonstration that semantic effects on orthographic processing can occur over and above any potential effects of phonological familiarity.

McKay et al. (2008) also examined the extent to which newly learned orthographic stimuli were treated similarly to real words by assessing whether naming latencies were sensitive to masked identity priming effects. They found that naming latencies for trained, but not untrained, novel words were decreased by masked identity primes relative to control primes. Furthermore, the size of the priming effect was equivalent to that typically seen in real words. On the basis of previous work demonstrating that masked identity priming effects are specific to words (and do not occur for nonwords) in lexical decision (Bowers, 2003; Forster et al., 2003; Forster, 1998; Forster & Davis, 1984) this was taken as evidence that newly learned items were treated similarly to real words. This in turn supported the idea that the results of this novel word

learning experiment inform us about orthographic processing in natural alphabetic languages.

However, as discussed in Chapter 1, the specificity of masked priming effects to words is not well established in naming, as it is in lexical decision. Masson and Isaak (1999) demonstrated equivalent masked identity priming effects for words and nonwords. In reviews of masked priming effects, both Forster et al. (2003) and Bowers (2003) concluded that in naming tasks, priming effects occur for both words and nonwords. Thus, it is debatable to what extent McKay et al.'s (2008) observations inform us about the similarity of newly learned and real word processing. If it could be shown that trained, but not untrained, novel words exhibit masked identity priming effects in old-new decision, this would provide stronger evidence that trained novel words are processed similarly to existing words.

The novel word learning experiments of Trudeau (2006) and McKay et al. (2008) seem to provide strong evidence for the role of semantics in orthographic processing, suggested by the triangle model (Harm & Seidenberg, 2004; Plaut et al., 1996). However, as mentioned in Chapter 1, there are several issues that deserve consideration. First, McKay et al. did not compare semantic vs. lexical phonology pre-exposure to a no pre-exposure baseline. It is therefore possible that phonological familiarity also supports orthographic processing, albeit to a lesser extent than semantic familiarity. Furthermore, both McKay et al. and Trudeau manipulated semantics within subject and it is possible that this introduced artefactual differences between experimental conditions. To explain further, in Trudeau's experiment, high imageability

definitions were probably more memorable than low imageability definitions. As participants learned both types of definition simultaneously, they may have paid greater attention to high imageability definitions which could have reduced the amount of information extracted about low imageability definitions. Similarly, in McKay et al.'s experiment, the task of learning the definitions in the semantic condition was interspersed with listening to and repeating the items in the lexical phonology condition. An efficient strategy would have been to pay minimal attention during the listen-repeat task, leaving full attention for the demanding semantic learning task. This may have reduced the benefit of lexical phonology pre-exposure, underestimating the role of phonological familiarity in orthographic processing and overestimating the importance of semantic familiarity.

Experiment 1

Experiment 1 sought to clarify the role of semantics in orthographic processing, suggested by the novel word learning experiments reported by Trudeau (2006) and McKay et al. (2008). Eighty adults were trained to read novel words written in English orthography. Each participant learned 12 items with consistent pronunciations and 12 items with inconsistent pronunciations. Prior to learning to read the novel items, participants received one of four pre-exposure conditions: 1) high imageability, or 2) low imageability definitions, both of which involved listening to the novel items and learning an associated novel definition and constituted *semantic* pre-exposure, 3) *lexical phonology* pre-exposure which involved listening to the sounds of the novel items, 4) a

baseline *no pre-exposure* condition. These conditions were manipulated between subjects. Outcome measures included reading accuracy during training and in a naming post-test, and naming and old-new decision latencies.

The following predictions were made: first, lexical phonology pre-exposure should provide greater benefit than no pre-exposure. Second, if semantic effects are genuine, definitions pre-exposure should provide greater benefit than lexical phonology pre-exposure. Note that this is in a between subjects design where learning definitions in the semantic conditions could not interfere with attending to item sounds in the lexical phonology condition. Third, if semantic pre-exposure is beneficial, this should be more pronounced in the high imageability definitions condition than in the low imageability definitions condition. It was also predicted that pre-exposure effects would be restricted to inconsistent items, both during training and in the naming post-test. However, in old-new decision, pre-exposure effects were expected across items, given that existing research has not demonstrated a semantics by consistency interaction in lexical decision. Note also that consistent items were expected to outperform inconsistent items across outcome measures.

To explore the extent to which trained items were treated in a similar way to real words, a masked priming manipulation was included in the old-new decision task with the prediction that identity priming (relative to control priming) would decrease response latencies to trained, but not untrained, novel words. We did not examine masked priming in the naming post-test because, as outlined earlier, there is no strong

evidence from natural language research that masked priming effects in naming are dependent on item lexicality.

Method

Participants

Eighty adults aged 18-30 years took part. All participants had English as a first language and had normal, or corrected to normal, hearing and vision.

Materials

Orthographic stimuli. A set of 24 novel word training items were adapted from those used by Trudeau (2006). Each item could be pronounced in two ways: a consistent pronunciation, pronounced in the same way as most English words with same orthographic body (e.g., RINT pronounced to rhyme with MINT), and an inconsistent pronunciation, pronounced in the same way as only one or two English words with the same orthographic body (e.g., STOME pronounced to rhyme with SOME). The distinction between consistent and inconsistent items was therefore relative, whereas in McKay et al.'s (2008) experiment it was all-or-nothing: inconsistent items had a pronunciation that never occurs in English for words with the same orthographic body. As discussed in Chapter 1, in English orthography, consistency is a graded variable. Thus, using Trudeau's stimuli approximates normal reading processes more closely.

All training stimuli had the following properties: each item began and ended with a consonant or consonant cluster and had a central vowel; the source of the difference between the consistent and inconsistent pronunciation was the vowel; each word body

had only two alternative pronunciations in English, the consistent pronunciation occurring in at least twice as many words in English as the inconsistent pronunciation; and each item had a different word body spelling. These lexical statistics were taken from the ARC nonword database (Rastle, Harrington, & Coltheart, 2002). Training items and their alternative pronunciations are given in Appendix 2.A. Stimuli were presented on a computer monitor in uppercase arial font, size 24.

Forty-eight further nonword stimuli were created for use in old-new decision. Twenty-four of these served as untrained distractor items and 24 as control primes. Both of these sets were individually matched for onset to the training items and began and ended with a consonant or consonant cluster and had a central vowel.

Phonological stimuli. The pronunciations of the training items were recorded by a female speaker and digitised at a sampling rate of 44Hz. All auditory stimuli were presented through headphones.

Semantic stimuli. Two sets of 24 novel definitions were selected from those devised by Trudeau (2006). Definitions were designed such that they did not closely resemble any real object or concept. One set were high in imageability and the other low in imageability (see Appendix 2.B). High imageability definitions described physical objects such as “sausage made with fish meat” or “an open-topped cargo container for bulk shipping”. Low imageability definitions described relatively abstract concepts such as “seeing similarities between facts” or “the ability to tell time without a clock”. The particular 24 definitions chosen were those that were rated highest and lowest in imageability by 16 individuals in Trudeau’s experiment. There was a highly significant

difference between the imageability ratings of these two sets, $t(23) = 7.68, p < .001$. The two sets of definitions were matched for number of words and each was presented on the computer screen in lowercase arial font, size 24.

Procedure

All 80 participants completed the following three phase procedure: first, a pre-exposure phase; second, an orthography learning phase; and third, an old-new decision task which involved discriminating trained from untrained items. A subset of 40 participants completed a naming post-test which took place directly after the orthography learning phase.

Pre-exposure. 20 participants took part in each of the following types of pre-exposure: 1) High imageability definitions, 2) Low imageability definitions, 3) Lexical phonology, 4) None. In definitions (semantic) conditions, a definition appeared on the screen in written form while the phonological form of the associated training item played through headphones. Participants were instructed to try to remember the association between the definitions and the items. All 24 items were presented in a random order and presentation was self-paced. This process was then repeated three more times. A fifth and final pre-exposure trial assessed semantic learning. Participants were presented with two definitions, one at the top of the screen and one at the bottom, and were asked to decide which one matched a spoken training item. Each item appeared once with its definition target in a randomised order. The pairing and spatial location of targets and distractors was pseudorandomised but was the same for each participant. No feedback was given.

In the lexical phonology condition, participants listened to the phonological forms of the 24 training items five times. As in the semantic conditions, presentation was self-paced and randomised within each block of 24 items. The no pre-exposure condition was a control condition in which participants did not have any exposure to the items before starting the orthography learning phase.

Orthography learning phase. The consistency manipulation was counterbalanced such that 40 participants (10 from each pre-exposure condition) learned the consistent pronunciation of 12 items and the inconsistent pronunciation of the other 12 items, and the remaining 40 participants received the reverse assignment. Participants were first exposed to the orthographic forms of the training items and their pronunciations by viewing each item, listening to its pronunciation, and repeating it once. On each trial, the item remained on the screen until the participant repeated it. Presentation was randomised. Participants then moved into the learning phase in which they viewed each training item and attempted to read it. Response times were not restricted. On each trial, the correct pronunciation was provided as feedback. After each item had been attempted once, the process was repeated twice more. Presentation was randomised within each block of 24 items. The experimenter recorded whether responses were correct throughout the exposure and learning phase.

Naming post-test (subset of 40 participants). Only 10 participants from each pre-exposure condition completed the naming post-test due to restricted availability of the voice-sensitive microphone (voice key). The task assessed the speed and accuracy with which participants could read the 24 training items. An item appeared and remained on

screen until the participant had read it out loud into the microphone. The next item was then presented. Participants were instructed to read each item as quickly and accurately as possible. The voice key was connected to a serial response box which recorded RTs. Actual responses were recorded using a desktop microphone to be scored offline for accuracy.

Old-new decision. This assessed participants' ability to discriminate trained items from untrained but similar items. It also investigated whether items were sensitive to masked identity priming. Targets were the 24 training items and distractors were 24 additional nonwords, as described in the Materials. Each item was presented twice, once with an identity prime and once with a control prime (all primes were presented in lowercase arial font, size 24). Control primes were the same for targets and distractors and consisted of 24 additional nonwords, again as described in the Materials. Each trial proceeded as follows; a mask (####) was presented for 450ms, a blank screen appeared for 50ms, a prime was presented for 50ms, followed by the item which remained on the screen until a response was made. Participants were instructed to press "z" if they thought the item was one they had learned and "m" if they thought it was one they had not learned. Response times were not restricted but participants were asked to make their decisions as quickly and accurately as possible. Item presentation was randomised and accuracy and RT data were recorded.

Results

Semantic learning

Participants in both the low and high imageability conditions successfully learned the association between the phonological forms of the items and their definitions. The proportion of items correctly selected in the semantic choice task was .95 ($SD = .21$) in the low imageability condition and .98 ($SD = .15$) in the high imageability condition.

Training

Figure 2.1 shows reading accuracy during training for inconsistent and consistent items, following the four types of pre-exposure. Accuracy was higher for consistent than inconsistent items throughout training. By the end of training (Block 3), participants read 70-80% of inconsistent items and 90-100% of consistent items correctly. Throughout training, performance was better in the semantic (high and low imageability definitions) conditions than in the no pre-exposure condition. This effect was more pronounced for inconsistent items and earlier in training. Accuracy was also higher in the lexical phonology than the no pre-exposure condition in training block 1 (although not as high as in the semantic conditions) for both consistent and inconsistent items. This effect dissipated in later blocks.

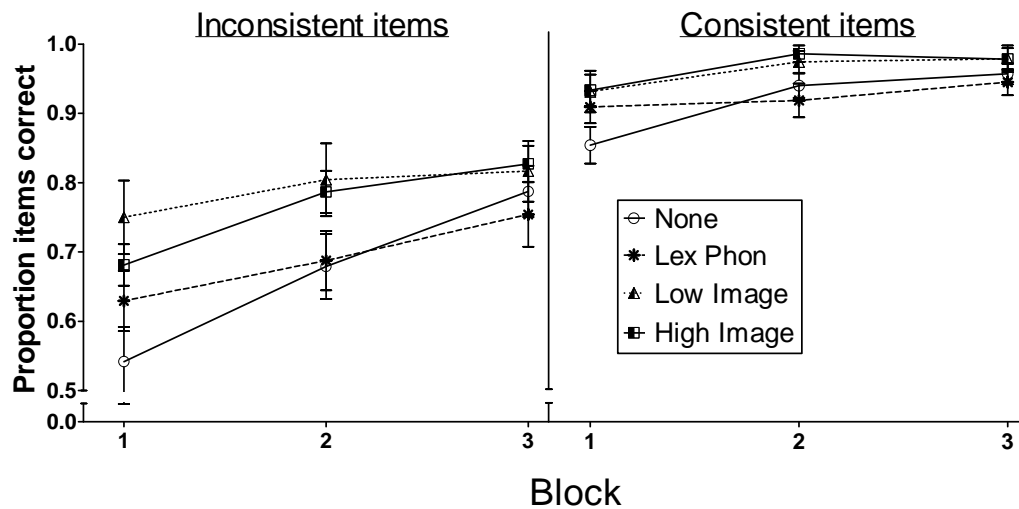


Figure 2.1. Accuracy during training as a function of pre-exposure and consistency. Vertical lines depict standard errors of subject means (SE).

The effects of pre-exposure (none vs. lexical phonology vs. low imageability vs. high imageability) and consistency (consistent vs. inconsistent) during training were assessed by conducting an analysis of variance on the proportion of items read correctly in each block of training (B1 vs. B2 vs. B3). In the subjects analysis (F_p), pre-exposure was a between subjects factor and consistency and block were within subject factors. In the items analysis (F_i), all factors were within item. Note that consistency was a repeated measures factor in the items analysis because each item had a consistent pronunciation for half the participants and an inconsistent pronunciation for the other half of the participants. In all analyses of variance reported in this and subsequent chapters, degrees of freedom were corrected using Greenhouse Geisser estimates if Mauchley's test indicated that the assumption of sphericity had been violated.

There was a main effect of pre-exposure, $F_p(3, 76) = 3.14, p < .05, \eta_p^2 = .11, F_i(3, 69) = 8.49, p < .001, \eta_p^2 = .27$. Across subjects, the only pairwise comparison to approach

significance ($p < .1$) was the advantage of the low imageability over the no pre-exposure condition, with the high imageability and lexical phonology conditions falling in between. The facilitative effect of semantic pre-exposure was clearer in pairwise comparisons across items: accuracy was higher in both the semantic conditions than in the no pre-exposure condition. In addition, accuracy in the low imageability condition was also higher than in the lexical phonology condition.

Accuracy improved in each subsequent block of training, $F_p(2, 156) = 61.70, p < .001, \eta_p^2 = .45, F_i(2, 46) = 71.80, p < .001, \eta_p^2 = .76$. The effect of block interacted with pre-exposure, $F_p(6, 152) = 4.24, p = .001, \eta_p^2 = .14, F_i(3.97, 91.19) = 4.15, p < .01, \eta_p^2 = .15$. Pairwise comparisons demonstrated that in the no pre-exposure and lexical phonology conditions, accuracy increased in each block of training, whereas in the semantic conditions, accuracy did not increase between blocks 2 and 3. Regarding the effect of pre-exposure in each block of training; in block 1, pairwise comparisons across both subjects and items demonstrated that accuracy was higher in the semantic conditions than in the no pre-exposure condition. In the low imageability condition accuracy was also higher than in the lexical phonology condition. In blocks 2 and 3, none of the pairwise comparisons across subjects reached significance. However, across items, accuracy in block 2 was higher in the semantic conditions than in the lexical phonology and no pre-exposure conditions. In block 3, accuracy was higher in the low imageability condition than in the lexical phonology condition, with accuracy in the high imageability and no pre-exposure conditions falling in between.

Performance was better on consistent than inconsistent items, $F_p(1, 76) = 108.92, p < .001, \eta_p^2 = .59, F_i(1, 23) = 30.73, p < .001, \eta_p^2 = .57$. The interaction between pre-exposure and consistency was marginal by-items, $F_i(3, 69) = 2.35, p = .08, \eta_p^2 = .09$, but non-significant by-subjects, $F_p < 1$. Pairwise comparisons across items demonstrated that the effect of pre-exposure was significant for inconsistent but not consistent items. There was also an interaction between block and consistency, $F_p(2, 152) = 12.53, p < .001, \eta_p^2 = .14, F_i(2, 46) = 16.13, p < .001, \eta_p^2 = .41$, reflecting the fact that for inconsistent items, accuracy improved in each subsequent block of training, whereas for consistent items, accuracy did not improve between blocks 2 and 3. The three-way interaction between pre-exposure, consistency, and block was not significant, $F_p(6, 152) = 1.10, ns, F_i(4.16, 95.56) = 1.03, ns$.

Naming post-test (subset of 40 participants)

Trials on which RTs were less than 250 ms or greater than 2500 ms, or in which a voice key error occurred, were excluded from the analysis (9.3% trials). Figure 2.2 shows that accuracy was higher when naming consistent than inconsistent items. This advantage does not appear to be reflected in RTs. Pre-exposure did not affect accuracy when naming consistent items. However, for inconsistent items, accuracy was higher in the semantic conditions than in the no pre-exposure condition, with accuracy in the lexical phonology condition falling in between. The pre-exposure advantage was not reflected in RTs; in fact RTs were slowest in the high imageability condition.

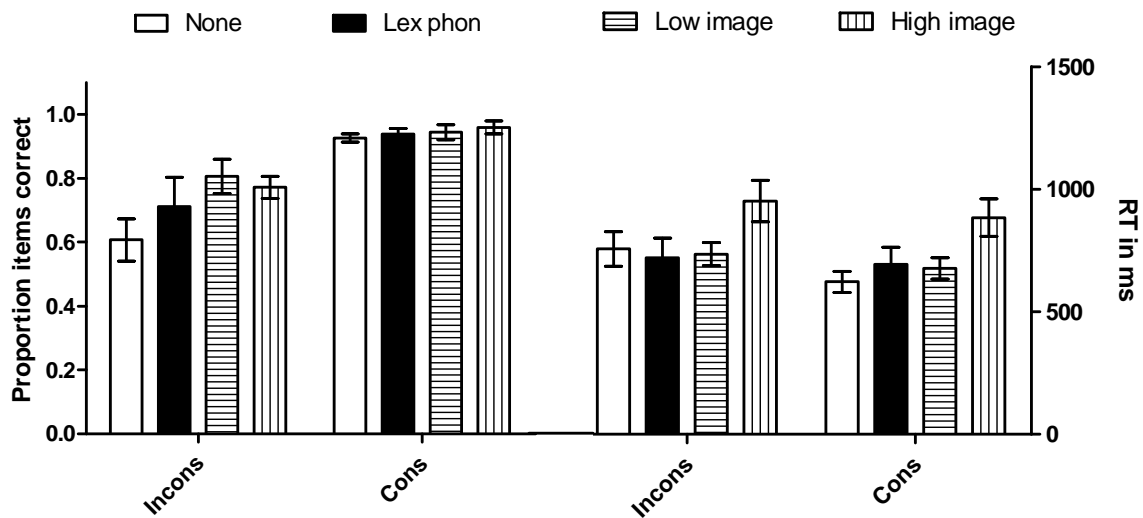


Figure 2.2. Accuracy and RTs (\pm SE) for trained item naming as a function of consistency and pre-exposure.

Two analyses of variance were conducted to examine the effects of pre-exposure and consistency on accuracy and mean RTs to correct items. In the accuracy analysis, there was a main effect of pre-exposure across items, $F_i(3, 69) = 6.90, p < .001, \eta_p^2 = .23$, but not subjects, $F_p(3, 36) = 1.81, ns$. Pairwise comparisons across items demonstrated that accuracy was higher in the semantic conditions than in the no pre-exposure condition, with accuracy in the lexical phonology condition falling in between. In the latency analysis, there was also a main effect of pre-exposure, $F_p(3, 36) = 2.86, p = .05, \eta_p^2 = .19, F_i(3, 69) = 27.39, p < .001, \eta_p^2 = .54$, but this did not mirror that observed for accuracy. Instead, RTs were longest in the high imageability condition.

Accuracy was higher for consistent than inconsistent items, $F_p(1, 36) = 49.72, p < .001, \eta_p^2 = .58, F_i(1, 23) = 25.96, p < .001, \eta_p^2 = .53$. The interaction between consistency and pre-exposure was significant by-items, $F_i(3, 69) = 3.32, p < .05, \eta_p^2 = .13$, but not by-

subjects, $F_p(3, 36) = 1.54$, *ns*. Pairwise comparisons across items demonstrated that the effect of pre-exposure was only significant within inconsistent items. Although not obvious from Figure 2.2, RTs were also faster for consistent than inconsistent items, $F_p(1, 36) = 17.52$, $p < .001$, $\eta_p^2 = .33$, $F_i(1, 23) = 15.71$, $p = .001$, $\eta_p^2 = .41$. The interaction between pre-exposure and consistency was not significant in the RT analysis, $F_p(3, 36) = 1.80$, $p = .16$, $\eta_p^2 = .13$, $F_i(3, 69) = 1.28$, *ns*.

The long naming latencies in the high imageability condition were unexpected and it was possible that they were driven by a small number of participants or items. However, the box and whisker plots in Figure 2.3 suggest that this was not the case. The long RTs were not restricted to a few outlying participants or items, and both the median and lower quartile values were higher in the high imageability condition than in all other conditions. Reasons for the long RTs in the high imageability condition will be examined in the Discussion.

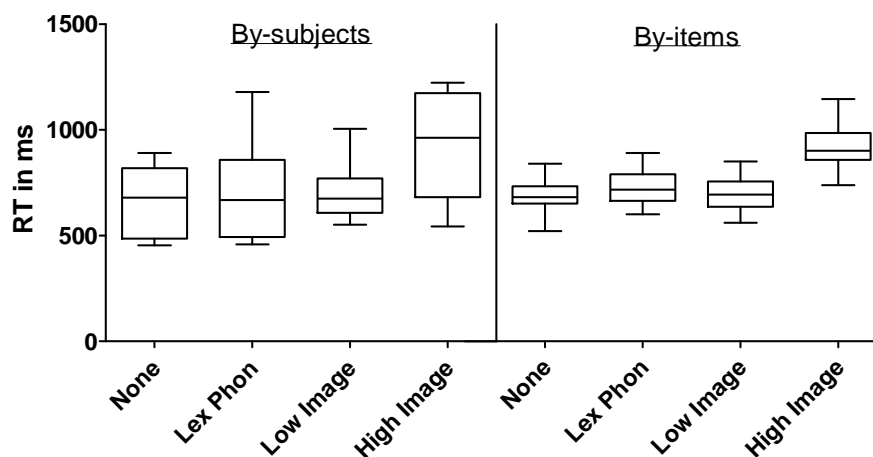


Figure 2.3. Subjects and Items box and whisker plot to show median, interquartile range, and range of naming RTs as a function of pre-exposure.

Old-new decision

Trials with RTs less than 250 ms or greater than 2500 ms were excluded (1.4%). Discrimination in all pre-exposure conditions was highly accurate ($\geq .95$ for trained items, $\geq .96$ for untrained items). Analyses of error data were therefore not carried out.

Masked priming effects. One purpose of the old-new decision task was to establish whether trained items were sensitive to masked identity priming. An analysis of variance on mean RTs was conducted to examine the effects of training status (trained vs. untrained) and prime type (identity vs. control) as a function of pre-exposure (none vs. lexical phonology vs. low imageability vs. high imageability). RTs were faster to trained, $M = 724$, than untrained items, $M = 739$, although this effect was only marginal by-subjects, $F_p(1, 76) = 3.02$, $p = .09$, $\eta_p^2 = .04$, and non-significant by-items, $F_i(1, 46) = 1.59$, *ns*. RTs were faster to identity primed items, $M = 720$, than to control primed items, $M = 742$, $F_p(1, 76) = 24.09$, $p < .001$, $\eta_p^2 = .24$, $F_i(1, 46) = 21.28$, $p < .001$, $\eta_p^2 = .32$. The interaction between training status and prime type was also significant, $F_p(1, 76) = 3.82$, $p = .05$, $\eta_p^2 = .05$, $F_i(1, 46) = 4.54$, $p < .05$, $\eta_p^2 = .09$. Pairwise comparisons confirmed that the prime effect was significant in trained (identity $M = 707$, control $M = 740$) but not untrained items (identity $M = 733$, control $M = 745$).

The main effect of pre-exposure was non-significant by-subjects, $F_p(3, 76) = 1.41$, *ns*, but significant by-items, $F_i(3, 138) = 37.61$, $p < .001$, $\eta_p^2 = .45$. Pairwise comparisons across items indicated that RTs in the high imageability condition, $M = 773$, were slower than in all other conditions and that RTs in the low imageability condition, $M = 684$,

were faster than in all other conditions. RTs in the lexical phonology, $M = 723$, and no pre-exposure conditions, $M = 745$, did not differ from each other. The effect of pre-exposure did not interact with training status, $F_s < 1$, or prime type, $F_p(3, 76) = 1.49$, *ns*, $F_i(2.54, 116.64) = 1.08$, *ns*, and the three-way interaction between training status, prime type, and pre-exposure was also non-significant, $F_p(3, 76) = 1.69$, $p = .18$, $\eta_p^2 = .06$, $F_i(23, 138) = 1.76$, $p = .16$, $\eta_p^2 = .04$.

Pre-exposure and consistency effects. Figure 2.4 shows decision latencies to trained items as a function of consistency and pre-exposure. As shown by the previous analyses and mirroring the naming post-test data, RTs in the high imageability condition were slower than in all other conditions, particularly for inconsistent items. However, RTs were also fastest in the low imageability condition.

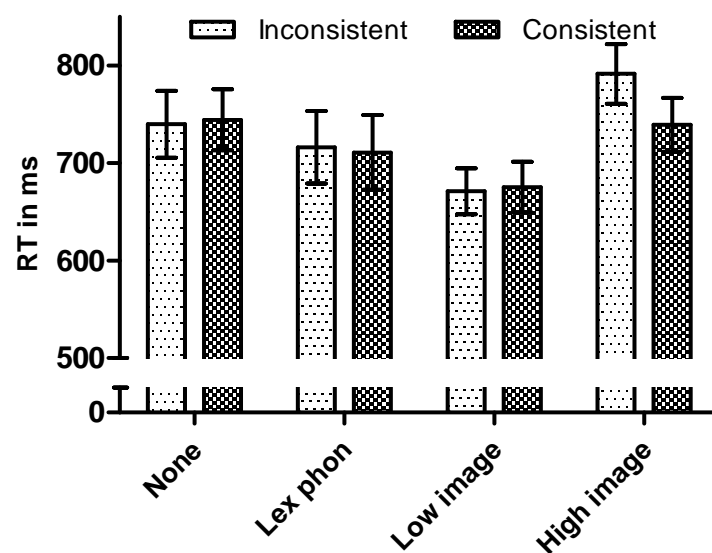


Figure 2.4. Old-new decision RTs (\pm SE) to trained items as a function of consistency and pre-exposure.

An analysis of variance on mean RTs to trained items was conducted to examine the effects of pre-exposure and consistency, collapsed across prime type. The main effect of pre-exposure was significant by-items, $F_i(3, 69) = 19.37, p < .001, \eta_p^2 = .46$, but not by-subjects, $F_p(3, 76) = 1.69, p = .18, \eta_p^2 = .06$. Pairwise comparisons across items confirmed that RTs in the high imageability condition were slower than those in the lexical phonology and low imageability conditions. In addition, RTs in the low imageability condition were faster than in all other conditions. The main effect of consistency was not significant, $F_p(1, 76) = 2.32, p = .13, \eta_p^2 = .03, F_i < 1$. However, consistency did interact with pre-exposure, $F_p(3, 76) = 2.85, p < .05, \eta_p^2 = .10, F_i(3, 69) = 2.69, p = .05, \eta_p^2 = .11$. Pairwise comparisons across both subjects and items demonstrated that in the high imageability condition only, RTs were faster to consistent than inconsistent items. In addition, pairwise comparisons across items revealed some differences in the effect of pre-exposure on consistent vs. inconsistent items. For consistent items, RTs in the low imageability condition were faster than in the no pre-exposure condition; RTs in the lexical phonology and high imageability conditions fell in between. For inconsistent items, RTs were slowest in the high imageability condition, and were also faster in the low imageability condition than in the no pre-exposure (but not the lexical phonology) condition.

Discussion

The aim of this experiment was to evaluate the prediction made by the triangle model of reading aloud (Harm & Seidenberg, 2004; Plaut et al., 1996) that semantic knowledge is particularly important when reading words with inconsistent spelling-sound mappings. This view is supported by data from acquired dyslexia (McKay et al., 2007; Patterson et al., 2006; Woollams et al., 2007), typical adult reading (Frost et al., 2005; Strain et al., 1995;2002; Woollams, 2005), and developmental research (Nation & Snowling, 2004; Ricketts et al., 2007). Lexical decision data from experiments with typical adults also suggest that semantic variables influence visual word recognition (Balota et al., 2004; Chumbley & Balota, 1984; Cortese & Khanna, 2007; Whaley, 1978). However, such natural language research has failed to acknowledge that the strength of a word's semantic and phonological representations is likely to be positively correlated. Thus, it is possible that at least part of the purported influence of semantics on orthographic processing is in fact driven by familiarity with an item's phonological form.

The current experiment separated the influence of phonological and semantic familiarity on orthographic processing by pre-exposing adults to the phonological form of novel words (lexical phonology) or to their phonological form plus either a high or low imageability definition (semantics). Adults then learned to read the novel words and the influence of phonological vs. semantic pre-exposure on subsequent orthographic processing was examined, with reference to a no pre-exposure baseline. Outcome measures included reading accuracy during training and in a naming post-test, and naming and old-new decision latencies.

Three specific predictions were made. First, phonological familiarity should facilitate reading aloud and old-new decision relative to no familiarity. Second, if semantic effects are genuinely semantic in nature, definitions pre-exposure should provide greater benefit than lexical phonology pre-exposure. Third, semantic effects should be more pronounced following pre-exposure to high, relative to low, imageability definitions. Pre-exposure effects were expected to be restricted to inconsistent items, except in old-new decision.

Regarding the first prediction, lexical phonology pre-exposure was beneficial in the first block of training but not in subsequent blocks. However, at post-test, lexical phonology pre-exposure did increase naming accuracy, particularly for inconsistent items. This provides some support for the idea that phonological familiarity aids reading aloud, particularly for items with somewhat atypical spelling-sound mappings. Turning to the second prediction, reading accuracy during training was higher in the semantic conditions than in the lexical phonology and no pre-exposure conditions. However, in the naming post-test, semantic pre-exposure did not provide greater benefit than lexical phonology pre-exposure. Facilitative effects of semantics on reading accuracy were restricted to inconsistent items both during training and at post-test. Overall, these findings provide some support for the view that semantics plays a role in reading aloud over and above the benefits conferred by phonological familiarity. However, they also suggest that phonological familiarity may account for some of the previously reported semantic effects on word naming. It should be noted that pre-exposure effects were generally significant by-items but either marginal or non-significant by-subjects. This is

probably due to the smaller number of participants in each condition (training: $n = 20$, naming post-test: $n = 10$), relative to the larger number of items ($n = 24$).

The third prediction, that high imageability definitions should provide greater benefit than low imageability definitions, was entirely unsupported. Reading accuracy was equivalent in the low and high imageability conditions both during training and at post-test. Furthermore, old-new decision and naming latencies were longer in the high imageability condition than in all other conditions. This unexpected effect was not due to outlying participants or items and raises the question of whether the observed benefit of semantic pre-exposure for reading accuracy was in fact due to the semantic content of the definitions. It seems likely that, in the semantic conditions, learning the association between the phonological forms of the items and their definitions focused participants' attention on listening to the items. In contrast, in the lexical phonology condition, although participants were instructed to listen to the items, no measure was put in place to ensure they did so. Attention to item sounds in the lexical phonology condition may therefore have been minimal. This leaves open the possibility that the benefit of semantic pre-exposure was not semantic in origin, but an artefact of increased attention to item sounds.

One finding that remains to be explained is why old-new decision and naming RTs were slowest in the high imageability condition. One possibility is that participants in the high imageability condition were actively trying to recall the definitions during naming and old-new decision and this slowed their responses. However, this is somewhat contradicted by the fact that responses were not slowed following pre-

exposure to low imageability definitions and the semantic recall task demonstrated that these were remembered as well as high imageability definitions. Furthermore, in the analysis of masked priming effects, participants in the high imageability condition were slower than participants in all other conditions, irrespective of whether they were responding to trained or untrained items. Untrained items did not have a definition to be recalled, therefore, participants should not be slowed on such items if active recall of definitions was responsible for the long response latencies in this group. A second possibility is that participants in the high imageability condition just happened to be slow. Although group differences were not driven by outlying participants or items, there is still a risk that pre-existing individual differences may be responsible for this finding, particularly given the relatively small participant numbers.

Evaluating these two possibilities would require more information about semantic learning in the low and high imageability conditions, for example, by testing explicit recall of definitions at the end of the experiment and/or by assessing the speed with which participants could remember definitions, perhaps using a more difficult four-alternative semantic choice task. To determine if there were group differences in basic processing speed, simple RTs on an independent task could also have been assessed. Another way to investigate reasons for this unexpected result would be to conduct a within subject version of the current experiment. In a within subject design, the slow responses in the high imageability condition should replicate if they were driven by the pre-exposure manipulation but not if they were due to underlying differences between participant groups. However, as discussed earlier, in a within subject design there is the

potential for pre-exposure conditions to interfere with one another, therefore such an experiment would come with its own possible confounds.

One reason for including the old-new decision task was to determine whether trained items were treated similarly to real words. In the current experiment, the mean decision latency for trained items was 723ms. This is fairly similar to the RTs reported by Stone et al. (774ms), Lacruz and Folk (620ms), and Balota et al. (680ms) for lexical decisions to English words. In typical lexical decision tasks, masked identity priming effects are generally observed for words but not nonwords (Bowers, 2003; Davis et al., 2008; Forster et al., 2003; Forster, 1998). Consistent with this, old-new decisions to trained items were faster following an identity prime than following a control prime and there was no prime effect for the untrained items. This effect was observed across pre-exposure conditions and supports the idea that trained items were treated in a similar way to real words.

A further point to note about old-new decision latencies is that they were largely insensitive to the consistency manipulation. This contrasts with research on English orthography demonstrating that consistent words elicit faster responses in lexical decision than inconsistent words (Balota et al., 2004; Cortese & Khanna, 2007; Lacruz & Folk, 2004; Stone et al., 1997). There are several differences between the current experiment and these studies which may account for this discrepancy. First, there was no frequency manipulation in the current experiment and Lacruz and Folk found that consistency had a greater influence on low frequency than on high frequency words. Although each training item was presented a relatively small number of times, it may be

that the recency with which they had been experienced negated any potential effects of consistency. A second possibility is that the consistency manipulation used in the current study was insufficient to influence old-new decision latencies. However, this seems unlikely because several studies on English words used a proportional measure of consistency that was similar to that used in the current experiment (Balota et al., 2004; Cortese & Khanna, 2007; Lacruz & Folk, 2004). Furthermore, McKay et al. (2008) also failed to obtain consistency effects for trained items in old-new decision, despite using a stronger all-or-nothing consistency manipulation. The third possibility is that the novel words in the current experiment were not controlled for feedback consistency. As described in Chapter 1, feedback consistency refers to whether an orthographic body has multiple possible spellings in English. Stone et al. (1997) demonstrated that feedforward consistency effects in lexical decision are more robust when feedback consistency is controlled.

Problems with the current experiment

The current experiment provided some evidence for a role for semantics in reading aloud. Pre-exposure to definitions increased reading accuracy for trained items with inconsistent pronunciations, relative to a no pre-exposure baseline, and to a lesser extent, relative to lexical phonology pre-exposure. However, semantic effects did not transfer to naming or discrimination latencies, and were not modulated by imageability in the predicted direction. It is possible that semantic effects were unclear due to methodological problems. Although a novel word learning method was employed to avoid some of the problems that plague natural language research, the use of English

orthography may mean that this was not as successful as intended. Some specific problems with the novel word learning method will now be discussed.

Using novel words meant that stimuli did not need to be selected from limited numbers of real English words. However, novel words were still assigned a consistent and an inconsistent pronunciation on the basis of English words. The inadequacy of this binary classification can be illustrated by comparing the novel words FREW and FLEIN. In the first block of training, the proportion of participants who read the inconsistent form of FREW correctly was .325, whereas for FLEIN it was .825. This is probably due to the fact that the inconsistent pronunciation of FREW (rhymes with SEW) goes against 26 English words ending in -EW which rhyme with BLEW, whereas, the inconsistent pronunciation of FLEIN (rhymes with STEIN) only goes against three English words ending in -EIN which rhyme with VEIN. As stated earlier, feedback consistency was also left to vary. If the degree of item consistency could be better controlled, performance might vary less across participants, increasing the clarity of any semantic effects.

A second problem with the classification of novel words as consistent or inconsistent is that neither the regularity of the grapheme-phoneme correspondences (GPCs) nor the influence of novel word onsets were considered. Although the consistency of the orthographic body accounts for a greater proportion of the variance in naming latencies than either of these other variables (Treiman et al., 1995), the results of the current experiment suggest that they nevertheless play a role. At a GPC level for example, participants found the inconsistent form of the novel word STOME (rhymed with SOME and COME) very difficult to learn; across the three training blocks

the mean accuracy for this item was .30, .38, .55. This may well be because the grapheme O is pronounced /ʌ/ in very few English words. At an onset-vowel level, the inconsistent version of the novel word NAFT was pronounced to rhyme with WAFT and in fact this pronunciation of the letter A typically only occurs in English words with a W onset. Again participants found this difficult to learn (mean accuracy in blocks 1-3, .25, .45, .53). As stated by Plaut et al. (1996, p. 59) "... if experimenters had to consider consistency across orthographic neighbourhoods at all possible levels, from individual graphemes up to the largest sub-word-sized chunks, their selection of stimulus words would be an even more agonizing process than it already is..." However, this does not remove the fact that such problems exist and influence performance.

The use of a novel word learning paradigm meant that all items could be learned simultaneously, controlling for AoA. However, the age at which we acquire sublexical spelling-sound units must also vary and could impact on learning. For example, it seems likely that we experience different pronunciations of the orthographic body -AVE (GAVE, WAVE, SAVE vs. HAVE) earlier than we experience different pronunciations of -ASTE (HASTE, PASTE, TASTE vs. CASTE). Such differences might change the nature of sublexical orthography-phonology representations and influence our ability to learn about such spelling-sound correspondences in the future. This illustrates a more general problem; when learning novel words in a familiar orthography, past experience with that orthography can impact on performance.

A further limitation is that because novel words were written in English orthography and pronunciations were analogous to English words, reading accuracy was

relatively high, even at the beginning of training and particularly for consistent items. This means that the current experiment did not investigate the contribution of semantic/phonological familiarity in earlier stages of learning to read. McKague, Pratt, and Johnston (2001) found that 6- to 7-year-old children benefited from prior phonological familiarity when learning to read novel words and that semantic familiarity provided no further benefit. However, as the consistency of the novel words was not manipulated, McKague et al.'s experiment does not address how familiarity with word meanings and sounds might differentially influence the early stages of learning to read consistent vs. inconsistent words. This issue remains poorly understood and the paradigm used in the current experiment cannot adequately address it.

The way forward

These problems raise concerns about the utility of novel word learning paradigms using a familiar orthography for investigating the role of semantics in orthographic learning. In subsequent experiments, an artificial orthography paradigm is developed to rectify these problems. Adults are exposed to novel words written in novel characters. They are not taught about individual character-sound correspondences, only whole-word pronunciations. This enables direct examination of the early stages of learning to read an alphabetic script. The consistency and frequency of the character-sound mappings are manipulated, making it possible to observe the process by which learners extract sub-word spelling-sound regularities. Experiments 2 and 3 focus on the influence of frequency and consistency at different stages of learning and Experiments 4 and 5 examine the influence of prior familiarity with phonological form vs. meaning.

CHAPTER 3: HOW DO WE LEARN SPELLING-SOUND MAPPINGS?

AN ARTIFICIAL ORTHOGRAPHY PARADIGM

A skilled reader of English is able to read and recognise words with both typical and atypical spelling-sound mappings, and generalize to novel words. This experiment used an artificial orthography paradigm to investigate how we learn to do so. Twelve adults were exposed to 36 novel words written in novel characters. The frequency and predictability of character-sound mappings was manipulated. Adults learned to read the novel orthography to at least 80% proficiency and could read generalization items, demonstrating extraction of character-sounds. Post-training, discrimination of trained from untrained items was above chance but laboured, suggesting that whole-item orthographic representations were weak. Learning and generalization were influenced by character-sound frequency and predictability, mirroring findings from natural languages.

Experiment 1 provided some support for the view that semantics influences reading aloud. However, semantic effects were not consistent across outcome measures and attention to item sounds may have been greater during semantic than lexical phonology pre-exposure. This leaves open the possibility that phonological rather than semantic familiarity supports orthographic processing. It was suggested that a more controlled paradigm might clarify semantic effects. One problem with the method used in Experiment 1 is that although participants learned to read novel words, they were

written in English orthography. Proficiency was therefore high from the beginning, and the experiment failed to address the question of how semantic and phonological familiarity influence *learning* to read. A further and equally important issue is that although research suggests that semantic influences on reading aloud vary according to word frequency and spelling-sound typicality, we in fact know very little about how sensitivity to these more basic variables develops. The aim of the current experiment was therefore to develop a more tightly controlled paradigm with which to investigate how the frequency and typicality of spelling-sound patterns influence orthographic learning. A semantic manipulation was not included in this experiment, but Experiments 4 and 5 return to the question of how word meanings influence orthographic learning.

Research on the influences of frequency and spelling-sound typicality on orthographic processing is first reviewed, framed in terms of how models of reading account for these effects. This provides a recap of the literature presented in Chapter 1, but also builds on this by exploring the developmental research in more detail. Overall, this research has informed us about what, but not how, the reading system learns. Next, problems with existing research are considered. Ways in which research in other domains has attempted to overcome similar problems are then outlined. Finally, the method developed for the current experiment is introduced: the artificial orthography learning paradigm.

Frequency and spelling-sound typicality effects in orthographic processing

How do we represent spelling-sound information?

The DRC (Coltheart et al., 2001) and triangle (Harm & Seidenberg, 2004; Plaut et al., 1996) models of reading aloud propose different accounts for how we are able to read both regular/consistent and irregular/inconsistent words whilst maintaining the ability to generalize to novel words. The DRC suggests that we possess a non-lexical route which represents context-independent rules for converting graphemes to phonemes, necessary for reading novel words, and a lexical route which represents whole-word mappings between orthography and phonology, necessary for reading irregular words. In contrast, in the triangle model, spelling-sound mappings for all words are represented across distributed connections between sets of units which code for orthography, phonology, and semantics.

These models make different predictions for human reading behaviour. According to the DRC, irregular words, which do not conform to grapheme-phoneme correspondence rules (GPCs), cause conflict between the lexical and non-lexical routes and should, therefore, be read more slowly than regular words, which conform to GPCs. The latency cost for irregular words is less pronounced when they are high in frequency. This is because the lexical route can generate their pronunciations quickly allowing the correct stored pronunciation to block the incorrect regularised pronunciation (generated by the non-lexical route). The DRC also predicts that novel words should be read using GPCs.

In contrast, the triangle model is not endowed with knowledge of GPCs and instead extracts information about the similarity, or consistency, between different words' spelling-sound mappings. Consistency and regularity are not always synonymous. For example, all graphemes in the words BEST and BONE are pronounced regularly. However, although all words ending in -EST rhyme with BEST, this is not true of all words ending in -ONE, e.g., GONE, DONE. The triangle model is sensitive to such differences and performs better when reading words like BEST than when reading words like BONE. As in the DRC, inconsistent words suffer less if they are high in frequency. This is because their spelling-sound mappings become strong enough to counteract the inhibition from similarly spelled but differently pronounced consistent words. When reading novel words, the triangle model does not rely on GPCs and instead pronounces graphemes in a way that is appropriate given the surrounding letter context. For example, the phoneme correspondence of the grapheme l is /l/, but when reading the novel word CHIND, the model pronounces it /tʃaɪnd/ to rhyme with FIND, KIND, and MIND (Harm & Seidenberg, 2004; Plaut et al., 1996).

As discussed in Chapter 1, human reading behaviour appears to provide stronger support for the triangle model than for the DRC. In particular, a body of work has shown that a graded measure of spelling-sound consistency accounts for more of the variance in naming latency than regularity (Cortese & Simpson, 2000; Jared, 2002; Jared et al., 1990; Treiman et al., 1995). In addition, Jared (2002) demonstrated that consistency, interacts with word frequency in its effects on reading aloud: inconsistent words suffer less of a naming latency cost when they are high frequency, and consistent words are

less influenced by frequency than inconsistent words. These findings question the extent to which our reading system represents explicit pronunciation rules and divides words into those that do and do not obey these rules, as suggested by the DRC model of reading aloud. Instead it supports the idea that spelling-sound knowledge is sensitive to context, as suggested by the triangle model.

Lexical decision and novel word reading tasks provide further support for the idea that we represent spelling-sound information in a graded fashion. Stone et al. (1997) found that consistency influences the speed with which we can discriminate words from nonwords, a finding that has now been replicated (Lacruz & Folk, 2004). Lacruz and Folk additionally demonstrated that consistency effects interact with frequency in lexical decision. Thus, even in a task in which no overt phonological output is required, the predictability and frequency of orthography-phonology mappings can influence orthographic processing. Treiman et al. (2003) found that adults applied context-sensitive consistency information when reading novel words. For example, DRANGE was more often read to rhyme with RANGE, MANGE, and CHANGE, whereas DRAND was more often read to rhyme with HAND, BLAND, and LAND. This shows that when pronouncing graphemes such as A, we apply our knowledge of how graphemes are pronounced in the context of other letters. This knowledge is gained through experience with reading words. Overall these findings support the idea that the same representations underlie our ability to read and recognise individual items and to generalize, as embodied in the triangle model.

How do we learn spelling-sound mappings?

The DRC is explicitly non-developmental, with Coltheart and colleagues avoiding learning issues on the grounds that "... unless the learning procedure itself is known to be psychologically real, it may not be able to learn what people learn..." (Coltheart et al., 2001, p. 216). This reduces this model's utility for examining questions about development. In contrast, the triangle model generates the explicit prediction that sensitivity to frequency and spelling-sound consistency should emerge through exposure to whole-word forms over the course of reading development. This prediction can to some extent be evaluated by considering the literature on children's reading development. Weekes, Castles, and Davies (2006) found that 8-year-old children were more accurate when reading consistent than inconsistent words, and that 10-year-old children were only influenced by consistency for low frequency words. This provides some evidence that sensitivity to the similarity and frequency of spelling-sound patterns develops fairly early on in the process of learning to read. Unfortunately, however, the words in this study were not controlled for GPC regularity, therefore the results do not provide unequivocal evidence for early development of context-sensitivity.

More convincing evidence is provided by studies examining children's generalization behaviour. Treiman et al. (1990) found that a higher proportion of 7- and 9-year-old children could read a nonword if it contained the same orthographic body as many real words, e.g., JOAL and SUG, than if few real words had the same orthographic body, e.g., JUL and SOAG, despite the fact that all nonwords could all be read using GPCs. A slightly different approach was taken by Stuart et al. (1999) who compared 7-

year-old children's reading of regular-consistent, irregular-consistent, and ambiguously-inconsistent nonwords. Regular-consistent nonwords contained an orthographic body that is pronounced according to GPCs in all real words, e.g., BISH or JACE. Irregular-consistent nonwords contained an orthographic body which is pronounced in the same way in the majority of real words but which is not pronounced according to GPCs, e.g., VOOK (GPC = /u/ vs. BOOK, TOOK, LOOK) or RAST (GPC = /æ/ vs. FAST, MAST, LAST). Ambiguously-inconsistent nonwords contained an orthographic body that is pronounced regularly in some words and irregularly in others, e.g., PONE (LONE, DONE, GONE) or FOVE (COVE, LOVE, PROVE). Results indicated that children used a greater number of irregular pronunciations for irregular-consistent nonwords than for the other two nonword types, demonstrating sensitivity to context. Similar findings of increased irregular pronunciations for irregular-consistent nonwords were reported by Treiman, Kessler, Zevin, Bick, and Davis (2006). These results provide convincing evidence that awareness of context-sensitive spelling-sound relationship begins to develop early.

In the studies described above there was also evidence for use of GPCs. Treiman et al. (1990) found that once the contribution of orthographic body frequency was partialled out, children's knowledge of the GPCs in the nonwords accounted for significant additional variance in naming accuracy. Stuart et al. (1999) found that the number of correct responses was higher for regular-consistent than irregular-consistent nonwords, and also that children were as likely to give regular pronunciations as consistent pronunciations for irregular-consistent nonwords. Overall these findings suggest that children develop sensitivity to many different types of spelling-sound

pattern in the early stages of learning to read. They also suggest that children use the knowledge developed through experience with individual words for generalization.

Developmental research therefore suggests that knowledge of context-sensitive spelling-sound patterns is acquired early in development. However, it is difficult to evaluate the extent to which this knowledge emerges through exposure vs. explicit instruction. No studies have directly investigated what can be learned through exposure. Thus, the extent to which we acquire item-specific and generalization knowledge about spelling-sound mappings just through exposure remains an open question.

Problems with existing research

One problem that researchers face when assessing how variables such as frequency, consistency, and regularity influence processing is how to measure these variables. GPCs and the consistency measure developed by Jared et al. (1990) were based on monosyllabic words in the Kucera and Francis 1967 corpus. Whilst such corpora provide a reasonable estimate of the words experienced by an average adult, they are only a proxy for an individual person's experience. This is particularly evident in children who have had a more limited and/or different experience with words. Masterson, Stuart, Dixon, and Lovejoy (2003) developed the Children's Printed Word Database to improve the lexical statistics used in developmental research. However, databases will never be able to capture individual variation in exposure. Furthermore, measures of spelling-sound consistency and regularity may not capture all the variance

that should be attributed to these variables. For example, Treiman et al. (2003) found that when reading novel words, adults also took into account the most frequent pronunciation of the onset plus vowel. For example, the novel word WAB was often pronounced to rhyme with SWAB, reflecting the fact in English, A is usually pronounced /ɒ/ when preceded by W, e.g., WANT, WATCH, WAN. This finding was later replicated in 6- to 9-year-old children (Treiman et al., 2006).

As a further example of this issue, the focus so far has been on feedforward consistency, defined as the predictability of a pronunciation given a particular spelling, but feedback consistency, the predictability of a spelling given a sound, also influences processing. Stone et al. (1997) showed that unless feedback consistency is controlled, feedforward consistency effects are unreliable in lexical decision. Similarly, Lacruz and Folk (2004) reported that controlling for feedback consistency was necessary for an interaction between frequency and feedforward consistency to be observed in lexical decision. Whilst it is clearly very difficult for researchers to take into account all the possible measures of spelling-sound consistency, studies that do not do so may underestimate the influence of such variables, and as a result, risk overestimating the influence of others.

A third problem is that much of the described research has compared groups of high vs. low frequency and consistent vs. inconsistent words when in fact these variables are of a graded nature. An example of the inadequacy of this binary classification was given in Experiment 1. Learners found the inconsistent pronunciation of the novel word FLEIN (rhymed with STEIN) far easier to learn than the inconsistent

pronunciation of FREW (rhymed with SEW) presumably because the pronunciation of many more English words goes against the latter (FLEW, BREW, GREW, etc.). Thus, in assigning consistent or inconsistent pronunciations to novel words rather than varying consistency in a more continuous manner, important information was lost about the graded influence of consistency on orthographic processing.

MacCallum et al. (2002) reviewed the problematic effects that can result from dichotomizing continuous variables. Although their examples were taken from studies of individual differences, the issues are just as important when considering item level variables such as frequency and consistency. Three of their points are of particular relevance to the present discussion. First, where researchers choose to place category boundaries may differ making it difficult to compare the results of different studies. For example, when investigating differences between consistent and inconsistent word reading, some researchers allow regularity to vary (e.g., Jared et al., 1990; Stone et al., 1997), whereas others also ensure that consistent words are also regular and inconsistent words are also exceptions. This makes it difficult to make direct comparisons between the results of different studies.

The second relevant issue identified by MacCallum et al. (2002) is that the practice of dichotomization forces variables into an orthogonal design when in fact they may be correlated. In a large scale regression analysis conducted by Balota et al. (2004), both objective and subjective frequency were significantly negatively correlated with orthographic body consistency. Studies in which items are selected to be high or low consistency and high or low frequency ignore this correlation. MacCallum et al.

presented evidence that this may cause analyses of variance to overestimate differences between group means. A third potential hazard occurs when variables have nonlinear effects on the measure of interest. For example, in Balota et al.'s study the nature of the relationship between word frequency and lexical decision latency was quadratic. MacCallum et al. demonstrated that if variables with nonlinear effects are dichotomized and then used as one of two or more factors in an analysis of variance, significant interactions can be obtained simply as a result of the nonlinearity in the effects of one or more of the variables. Taken together, these issues suggest that the results from many of the studies on the influence of frequency and consistency on orthographic processing may be unreliable.

A final problem with the existing research is that there are many other variables, besides frequency and spelling-sound predictability, which influence word processing. Failing to control for such variables may mean that the influence of the measures of interest is overestimated. For example, many studies match experimental word lists for initial phoneme, but Rastle and Davis (2002) demonstrated that matching the entire word onset is crucial when using a voice key to obtain naming latencies. Voice keys are often triggered by the onset of voicing rather than the actual point at which a response begins. Because some items with consonant cluster onsets have unvoiced consonants as the second phoneme (e.g., spat), the onset of voicing occurs, on average, later for such words. Thus, voice keys may be triggered later for words with cluster onsets than words with single phoneme onsets (e.g., sat). This means that, unless the entire onset is

matched, spurious differences between lists may be obtained simply through voice key insensitivity.

A higher level example of an important but often uncontrolled variable is the age at which a word enters the spoken vocabulary. Although age-of-acquisition (AoA) is correlated with word frequency, both variables have a unique influence on naming latencies in adults (Gilhooly & Watson, 1981). The effect of AoA is also apparent in children's reading. Weekes et al. (2006) found that for a set of words matched for frequency, orthographic body consistency influenced 10-year-old children's reading accuracy for late, but not early, acquired words. These findings suggest that studies which do not take AoA into account risk inflating the influence of other variables.

The preceding discussion has demonstrated that not only has existing research failed to answer the question of how we learn spelling-sound mappings, but also that it suffers from many methodological problems. New methods are therefore needed to overcome these problems in order to corroborate or question existing findings and provide information about learning.

Artificial language learning studies

Artificial language learning paradigms have been used by a growing number of researchers to explore issues in spoken language processing, some examples of which were given in Chapter 1 (Creel et al., 2006; Magnuson et al., 2003; Wonnacott et al., 2008). This methodology has the advantage of giving complete control over the input statistics. It is therefore a useful tool for separating the influences of correlated

variables and directly examining learning. Unfortunately, very few studies have examined learning of novel orthographies. This is key if we are to understand how lexical variables such as frequency and spelling-sound predictability influence *learning to read*.

Studies that have used novel scripts have tended to focus on learning individual symbol-sounds, rather than sequences of symbols forming words. For example, Byrne (1984) and Byrne and Carroll (1989), taught adults new symbols for consonants and examined whether they extracted phonetic feature information embedded in the symbols. Some recent neuroimaging studies have also looked at changes in brain activity pre- and post-training on previously unknown orthographic symbols (Callan et al., 2005; Hashimoto & Sakai, 2004). However, none of these studies examined learning of spelling-sound patterns. Two experiments that have done so were conducted by Bitan and Karni (2003; 2004). Adults were exposed to novel words written in a novel orthography and extraction of sub-word spelling-sound correspondences was examined. Results demonstrated that although participants were able to learn the training sets, they were unable to generalize their knowledge to novel forms written in the same orthography without explicit teaching in symbol-sound correspondences. These findings suggest that children's sensitivity to spelling-sound patterns must be acquired through direct teaching.

Importantly however, there are several problems with Bitan and Karni's (2003; 2004) methodology which mean that this might not be the case. First, unlike most alphabetic scripts in which phonemes tend to be represented by a cohesive symbol,

phonemes in these studies were represented by two or three separate symbols. Furthermore, these symbols were not unfamiliar to participants and were therefore unlikely candidates for letters. For example, the sound /p/ was written as /[^]*, /b/ as */ , and /n/ as [^]/[^], meaning that the word /pɒn/ was written as /[^]**/[^]/[^]. This complexity was used to minimise the impact of existing alphabetic knowledge. However, it is likely to have actively hindered the extraction of sub-word symbol-sound knowledge. Second, in natural languages, children are soon faced with an ever increasing set of words, whereas, Bitan and Karni taught adults only 6 to 12 words. Coupled with the intractable nature of the sub-word symbol-sound relationships, it is not surprising that learners adopted a whole-word rote learning strategy. Third, feedback during learning involved participants deciding whether their decoding attempts matched an English word translation, and making same-different judgements about pairs of stimuli. This learning process is very different to the corrected pronunciation attempts that typically characterise how children learn to read new words. Together, these factors may well have promoted learning that was very different to that seen when people learn a natural alphabetic script. Thus, whether learners can extract context-sensitive symbol-sound correspondences from an artificial orthography and use them to support generalization to new forms remains an open question.

Experiment 2

This experiment examined learning and generalization in an artificial orthography. Adults learned to read a set of novel words written in novel characters.

They were not taught about individual character sounds, just about whole-word pronunciations. The frequency and predictability of the character-sound mappings was systematically varied enabling examination of the influence of these factors on learning. Following training, participants were tested on their ability to discriminate trained from untrained orthographic forms, and on their ability to read a further set of novel items written in the same characters as the training set. This paradigm enabled complete control over exposure to the statistics of the language. This means that precision of and confidence in results should be greater than those obtained in natural language research. Importantly, it also provided an environment to directly examine learning.

Four main questions were explored. First, can learners extract sub-word spelling-sound mappings from exposure to whole-word orthographic forms (and their corresponding pronunciations) without explicit instruction, and use this to support generalization? Second, is knowledge extracted about spelling-sound mappings sensitive to context? Third, are learning and generalization influenced by the frequency and predictability of spelling-sound mappings present in the language environment? Finally, do learners develop whole-item orthographic representations and are these also sensitive to spelling-sound frequency and predictability?

Method

Participants

Twelve adults (5 males, 7 females) took part. They were all university students and their mean age was 20.2 years, $SD = 2.1$. All participants had English as a first language and had normal, or corrected to normal, hearing and vision.

Materials and Procedure

All participants completed the same three phase procedure: first, an exposure and learning phase; second, an old-new decision task which involved discriminating the orthographic forms of trained from untrained items; and third, a generalization task that asked them to read aloud a set of untrained items. The experimental procedure is summarised in Figure 3.1 and explained in more detail below. The entire experiment was conducted in a single session lasting between 30 and 45 minutes.

Exposure and learning phase. Two sets of 36 training items were constructed. Half the participants were exposed to one set and half to the other in order to minimise the impact of any idiosyncratic properties of the items. All training items were monosyllabic consonant-vowel-consonant novel words. The items in each set were constructed from 12 consonant and 6 vowel phonemes. The consonants were the same in both sets (/b/, /d/, /f/, /g/, /k/, /m/, /n/, /p/, /s/, /t/, /v/, /z/) but the vowels differed. Set one contained the vowels /ɛ/, /i/, /aɪ/, /ɒ/, /əʊ/, /u/, and set two /ɪ/, /i/, /aɪ/, /əʊ/, /ʌ/, /ʊ/. The written forms of both sets of training items were constructed from the same 16 characters. However, to minimise any effects of particular phonemes being easier to map to particular characters, two different randomly assigned character-phoneme mappings were used and half the participants who were exposed to each training set experienced one set of character-phoneme mappings and half the other, see Appendix 3.A.

In both training sets, consonant phonemes were represented by a single character. However, vowel character-phoneme mappings varied in consistency and

frequency. Two vowel characters in each set were *consistent* and pronounced in the same way in all items. The other two vowel characters were inconsistent and were pronounced one way when preceded by a particular consonant character (*inconsistent-conditioned* pronunciation) and in a different way when preceded by any of the other consonant characters (*inconsistent-unconditioned* pronunciation). This gave a total of four vowel characters corresponding to six vowel phonemes. In English orthography, vowel pronunciations are affected more often by final than initial consonants (Treiman et al., 1995) whereas in the artificial orthography, the reverse was true. This meant that learning about the conditional relationships between consonant and vowel pronunciations could not be based on existing knowledge of English.

This consistency manipulation was crossed with a frequency manipulation. One of the *consistent* characters was *high frequency*, occurring in eight items, and one *low frequency*, occurring in four items. One inconsistent character was pronounced in the *inconsistent-conditioned* way in eight items and the *inconsistent-unconditioned* way in four items, while the other inconsistent character was pronounced in the *inconsistent-conditioned* way in four items and the *inconsistent-unconditioned* way in eight items. It should be highlighted that this is a type frequency manipulation in which vowel types varied in the number of different words they occurred in. In contrast, the effect of frequency on processing of English words, as described in the earlier literature review, generally refers to token frequency, how often a particular word occurs in the language. The influences of type and token frequency are highly correlated in natural languages (Kessler, Treiman, & Mullenix, 2003), but the interplay between them represents an

important area for future research. However, varying type frequency was considered sufficient for examining learners' sensitivity to the frequency characteristics of the language.

Except for the two consonants that formed the onset for inconsistent-conditioned items, which necessarily occurred as often as the vowel phonemes they affected (i.e., four or eight times), consonants occurred approximately the same number of times in onset position across training items. Similarly, all consonants occurred approximately the same number of times in coda position. Full details of the consistency by frequency manipulation for one of the training sets is given in Appendix 3.B; their pronunciations are provided in Appendix 3.C, Table 3.C1. Participants viewed written stimuli on a monitor and heard spoken stimuli (recorded by a female speaker and digitised at a sampling rate of 44Hz) through headphones. They wore a microphone and were recorded throughout the experiment.

In the exposure phase, participants viewed each training item, listened to its pronunciation and repeated it once. The item remained on the screen until the participant had repeated it, and the experimenter had recorded whether their pronunciation was correct. Presentation was randomised. Participants then moved into the training phase in which they viewed each training item and attempted to read it. Response times were not restricted. Once participants had provided a response, the correct pronunciation was given as feedback and the experimenter recorded whether their attempt was correct. All items were presented in a randomised order. If less than 80% of items were read correctly, all items were attempted again. This criterion level

was chosen because pilot studies indicated that at this level, participants were able to competently complete post-tests.

Old-new decision. This task was completed directly after orthography training and assessed participants' ability to discriminate trained items from untrained but similar items. The task included 12 targets (trained items) and 12 distractors (untrained items). Trained items comprised equal numbers of high and low frequency consistent, inconsistent-conditioned and inconsistent-unconditioned items. Untrained distractors were the same items used to assess generalization (see below). An item was presented and participants pressed "z" if they judged it to be one they had learned and "m" if they judged it to be one they had not learned. Response times were not restricted but participants were instructed to make their decisions as quickly and accurately as possible. Accuracy and RTs were recorded.

Generalization. Two sets of 12 generalization items were created, one for each training set. Four items in each set contained consistent vowel characters and eight contained inconsistent vowel characters. Of the eight items containing inconsistent characters, four were preceded by a consonant character that should cause the vowel to take the inconsistent-conditioned pronunciation and four by a consonant character that should cause the vowel to take the inconsistent-unconditioned pronunciation. Half the test items contained consistent, inconsistent-conditioned and inconsistent-unconditioned vowels that were high frequency during training and half contained vowels that were low frequency during training. This enabled assessment of the impact of the frequency and consistency of vowels during training on generalization

performance. An example of one of the generalization sets is shown in Appendix 3.C, Table 3.C2. Participants attempted to read each of the generalization items. Response times were not restricted and no feedback was given. Items were scored correct if consonants were pronounced correctly and if vowels were pronounced according to their consonant context, i.e., given the inconsistent-conditioned or inconsistent-unconditioned pronunciation when appropriate.

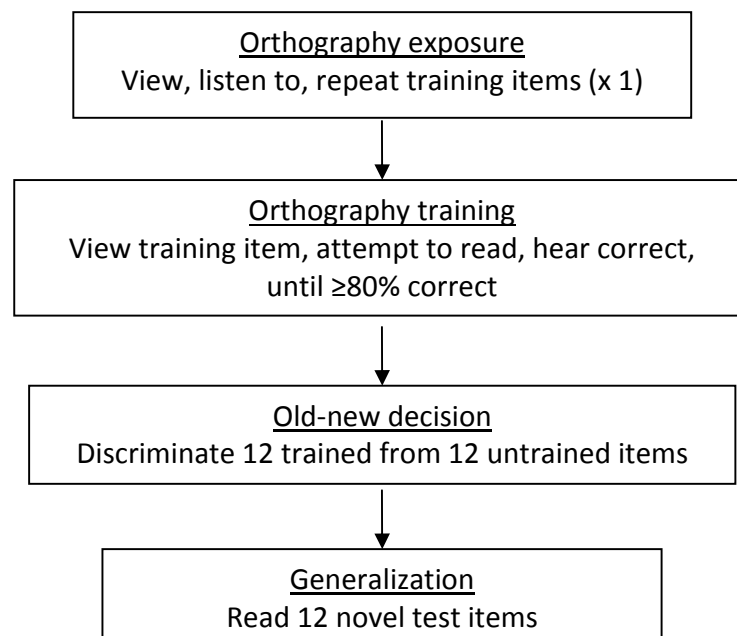


Figure 3.1. Summary of procedure for Experiment 2.

Debriefing. To determine whether participants had extracted explicit knowledge of the context-sensitive character-sound mappings, they were asked at the end of the experiment whether they had noticed that some characters had more than one pronunciation. They were then asked whether they had developed any rules that helped

them decide how to pronounce the words. The experimenter noted down the participants' responses and then provided them with a detailed explanation of how the language was constructed and the background and hypotheses to the experiment.

Results

Training

The maximum number of blocks required to achieve 80% accuracy was 7. For analyses of performance over the course of training, if a participant required less than 7 blocks to achieve criterion, accuracy in remaining blocks was assumed to be identical to that in the final block of training completed. For example, if a participant achieved criterion after 5 training blocks, it was assumed that the items they read correctly in Block 5 were also read correctly in Blocks 6 and 7.

Individual differences in learning. Participants varied in learning rate, as shown by the large standard deviations in Figure 3.2. The maximum number of blocks required to achieve 80% accuracy was 7 and the minimum was 2, with a mean of 5.0 ($SD = 2.0$). There was a substantial negative correlation between accuracy in Block 1 and the number of blocks to achieve 80% accuracy, $r = -.70$, $p = .01$, i.e., those who performed best at the beginning learned fastest.

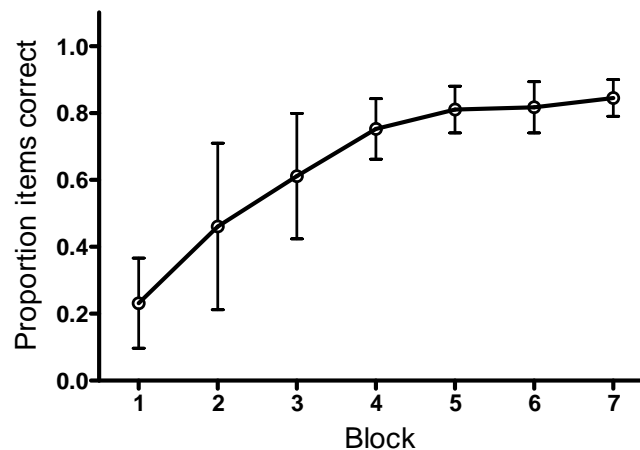


Figure 3.2. Proportion of items read correctly over the course of training. Vertical lines depict standard deviations of subject means (SD).

Frequency and consistency effects during training. Figure 3.3 shows reading accuracy over the course of training for items containing low and high frequency vowels as a function of vowel consistency. Within items containing low frequency vowels, the pattern of performance was similar throughout training: accuracy was highest for items containing consistent vowels, followed by items containing inconsistent-conditioned vowels, and lowest when items contained inconsistent-unconditioned vowels. Within items containing high frequency vowels, the pattern of performance was slightly more complicated. At the beginning of training, accuracy was highest when items contained inconsistent-conditioned vowels. However, towards the middle-end of training, performance on items containing consistent vowels increased to an equivalent level. Accuracy remained lowest on items containing inconsistent-unconditioned vowels throughout training. At the beginning of training, frequency facilitated performance on items containing consistent and inconsistent vowels; however, frequency only influenced performance on items containing inconsistent vowels by the end of training.

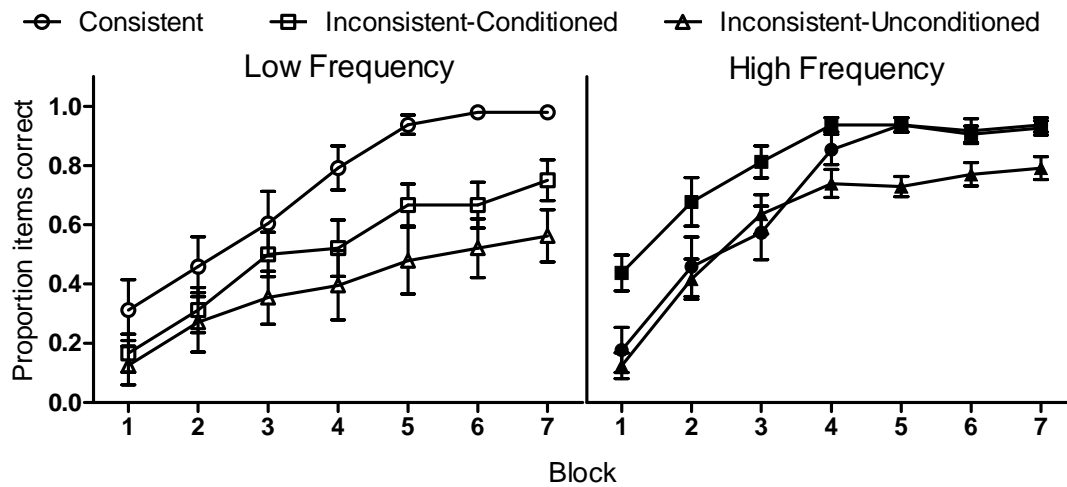


Figure 3.3. Proportion of items read correctly (\pm SE) over the course of training as a function of vowel frequency and consistency.

The effects of vowel frequency (low vs. high) and consistency (consistent vs. inconsistent-conditioned vs. inconsistent-unconditioned) were assessed by conducting an analysis of variance on the proportion of items read correctly in each block of training (Block 1 – Block 7). In the subjects analysis, frequency, consistency, and block were treated as within subject factors, whereas in the items analysis, frequency and consistency were between items factors and block was a within item factor.

Accuracy increased in each subsequent block of training, $F_p(1.79, 13.74) = 46.11$, $p < .001$, $\eta_p^2 = .81$, $F_i(4.04, 266.38) = 171.45$, $p < .001$, $\eta_p^2 = .72$. Accuracy was higher for items containing high frequency vowels than for those containing low frequency vowels, $F_p(1, 11) = 14.16$, $p < .01$, $\eta_p^2 = .56$, $F_i(1, 66) = 36.20$, $p < .001$, $\eta_p^2 = .35$, and for items containing consistent or inconsistent-conditioned vowels than for those containing

inconsistent-unconditioned vowels, $F_p(2, 22) = 14.68, p < .001, \eta_p^2 = .57, F_i(2, 66) = 23.94, p < .001, \eta_p^2 = .42$.

The interaction between frequency and block was not significant by-subjects, $F_p(2.77, 30.43) = 2.23, p = .11, \eta_p^2 = .17$, but was by-items $F_i(4.04, 266.38) = 4.17, p < .01, \eta_p^2 = .06$. Pairwise comparisons across items demonstrated that accuracy was higher for items containing high than low frequency vowels in all but the first block of training. Furthermore, while accuracy for items containing low frequency vowels continued to improve throughout training, it had reached asymptote by block 4 for items containing high frequency vowels. There was also an interaction between consistency and block, $F_p(12, 132) = 3.83, p < .001, \eta_p^2 = .26, F_i(12, 396) = 3.91, p < .001, \eta_p^2 = .11$. In earlier training blocks, accuracy for items containing consistent and inconsistent-conditioned vowels did not differ and was higher than for items containing inconsistent-unconditioned vowels. In contrast, accuracy in later training blocks was higher for items containing consistent vowels than for items containing inconsistent-conditioned vowels, which was in turn higher than for items containing inconsistent-unconditioned vowels. In addition, accuracy for consistent and inconsistent-conditioned items did not improve after block 4, whereas for inconsistent-unconditioned items it continued to increase until the end of training.

There was also an interaction between frequency and consistency, $F_p(2, 22) = 8.19, p < .01, \eta_p^2 = .43, F_i(2, 66) = 13.49, p < .001, \eta_p^2 = .29$. The facilitative effect of frequency was only significant within inconsistent and not consistent items. In addition,

for items containing low frequency vowels, accuracy was highest for those containing consistent vowels, followed by those containing inconsistent-conditioned vowels and was lowest for items containing inconsistent-unconditioned vowels. In contrast, for items containing high frequency vowels, accuracy was highest for inconsistent-conditioned items, followed by consistent items, and was lowest for inconsistent-unconditioned items. The three-way interaction between block, frequency and consistency was not significant, $F_p < 1$, $F_i(12, 396) = 1.15$, *ns*.

Frequency and consistency effects at the end of training. Figure 3.4 shows that, by the end of training, accuracy was higher for consistent than inconsistent items, particularly when they contained low frequency vowels. Accuracy was also higher for items containing high frequency vowels than for those containing low frequency vowels, particularly when vowels were inconsistent. The effects of vowel frequency and consistency at the end of training were assessed by conducting an analysis of variance on the proportion of items read correctly in the block in which at least 80% accuracy was achieved. It should be remembered that this constituted different blocks for different participants. In the subjects analysis, frequency and consistency were treated as within subject factors and in the items analysis they were treated as between items factors. These designs were also used to analyse the old-new decision and generalization data reported later in the Results.

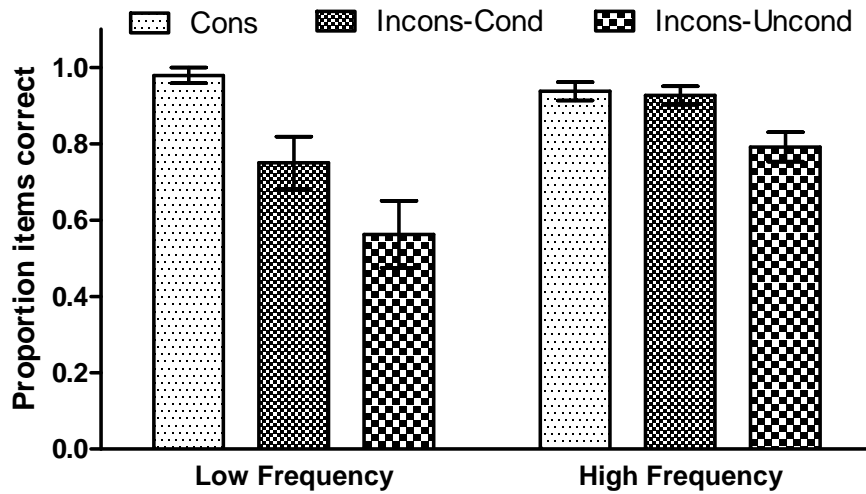


Figure 3.4. Proportion of items correct (\pm SE) in the block at which 80% accuracy was achieved, as a function of vowel frequency and consistency.

Performance was better for items containing high frequency vowels than for items containing low frequency vowels, $F_p(1, 11) = 7.07, p < .05, \eta_p^2 = .39, F_i(1, 66) = 10.13, p < .01, \eta_p^2 = .13$. Accuracy for consistent items was higher than for inconsistent-conditioned items, which was in turn higher than for inconsistent-unconditioned items, $F_p(2, 22) = 18.29, p < .001, \eta_p^2 = .62, F_i(2, 66) = 18.21, p < .001, \eta_p^2 = .36$. These main effects were qualified by a significant interaction, $F_p(2, 22) = 3.48, p = .05, \eta_p^2 = .24, F_i(2, 66) = 4.72, p = .01, \eta_p^2 = .13$. The facilitative effect of frequency was only significant in items containing inconsistent, and not consistent, vowels. Within items containing low frequency vowels, the effect of consistency was as reported for the main effect. However, within items containing high frequency vowels, accuracy was equally high for consistent and inconsistent-conditioned items and both were higher in accuracy than inconsistent-unconditioned items.

To investigate the types of errors made, responses were re-scored such that the alternative pronunciation of inconsistent vowel characters was also accepted as correct. When rescored in this way, the proportion of trained items read correctly in the final training block was .96 ($SD = .03$) demonstrating that the majority of errors resulted from using the alternative pronunciation for inconsistent vowel characters.

Old-new decision

Trials with RTs more than two standard deviations away from the mean for that participant were excluded from the analysis (2.8%). One sample t-tests indicated that old-new decision accuracy was above chance for trained, $t_p(11) = 5.90$, $p < .001$, but not untrained items, $t_p(11) < 1$. It was therefore possible that the above chance performance on trained items was driven by a YES response bias. To determine if this was the case, the proportion of hits (YES responses to trained item targets) and false alarms (YES responses to untrained item distractors) for each participant were transformed to z-scores. $z(\text{false alarms})$ was then subtracted from $z(\text{hits})$ to give a d' statistic for each participant. A one sample t-test indicated that d' was significantly different from 0, $t_p(11) = 2.66$, $p < .05$, demonstrating that although responses were subject to a yes bias, this was not entirely responsible for the above chance performance on trained items. It was therefore valid to conduct further analyses on responses to trained item targets.

Accuracy (SD in parentheses) was higher for trained, $M = .74$ (.14), than untrained items, $M = .44$ (.23), $t_p(11) = 3.60$, $p < .01$. RTs did not differ according to training status (trained $M = 2693$ (952), untrained $M = 3038$ (1333), $t_p(11) = 1.61$, $p = .13$). The effects of vowel frequency and consistency on old-new decision performance

were assessed by conducting two analyses of variance, one with proportion of trained items correct as the dependent variable and one using mean RTs to correct trained items. There were no significant main effects or interactions in either of these analyses, full details of which are provided in Appendix 3.D.

Generalization

Responses were scored as correct if consonants were pronounced correctly and if vowels were given the conditioned or unconditioned pronunciation when this was demanded by the preceding consonant. Generalization performance was fairly high; the mean proportion of items read correctly was .73 (.14). An analysis of variance was conducted to examine the effects of vowel frequency and consistency during training on the proportion of generalization items read correctly. The results of this analysis are summarised in Figure 3.5.

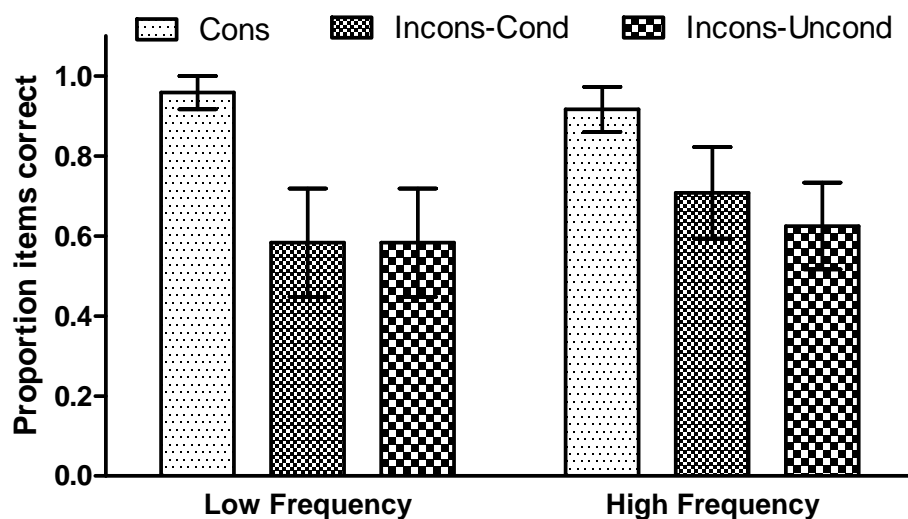


Figure 3.5. Proportion of generalization items correct (\pm SE) as a function of vowel frequency and consistency during training.

Vowel frequency did not influence generalization, $F_s < 1$. However, accuracy was higher for generalization items containing consistent vowels than for those containing inconsistent vowels, $F_p(1.26, 13.81) = 4.10$, $p = .06$, $\eta_p^2 = .27$, $F_i(2, 18) = 8.88$, $p < .01$, $\eta_p^2 = .50$. The interaction between frequency and consistency was not significant, $F_s < 1$.

As with the end of training data, responses were then rescored such that the alternative pronunciation of inconsistent vowel characters was also accepted as correct. When rescored in this way, the mean proportion of items read correctly was .96 (.08) demonstrating that the majority of errors resulted from using the alternative pronunciation of inconsistent vowel characters. The possibility was then considered that learners might have adopted a strategy of using only the high frequency pronunciation of the inconsistent vowel characters in generalization. However, when data were rescored such that only the high frequency pronunciation of the inconsistent vowel characters was accepted as correct, accuracy on items containing inconsistent vowels was in fact lower overall, $M = .46$ (.27), than when only appropriate context-dependent pronunciations were scored as correct, $M = .63$ (.34).

Discussion

The triangle model of reading aloud (Harm & Seidenberg, 2004; Plaut et al., 1996) suggests that sensitivity to the frequency and predictability of spelling-sound mappings should emerge through exposure to the orthographic forms of whole words and their pronunciations. Data from orthographic processing in typical adults (Jared, 2002; Jared et al., 1990; Lacruz & Folk, 2004) supports the view that our knowledge of

spelling-sound patterns is context-sensitive and influenced by frequency, and also suggests that this knowledge is used to support generalization to novel words (Treiman et al., 2003). Developmental evidence demonstrates that context-sensitive spelling-sound knowledge is acquired relatively early in the process of learning to read (Treiman et al., 2006; Weekes et al., 2006). However, existing studies have not been able to assess the extent to which this knowledge is acquired implicitly through experience with the language vs. explicitly through teaching of spelling-sound correspondence rules. In addition, there are many methodological issues with the way research to date has been conducted and these undermine confidence in results from such studies. In the current experiment, an artificial orthography learning paradigm was designed to overcome these methodological problems and directly examine the learning process by which spelling-sound knowledge is acquired.

Adults learned to read 36 novel CVC words written in novel characters. The vowels in these novel words varied in the consistency and frequency of their character-phoneme mappings. Two vowel characters were *consistent* and always took the same pronunciation, and two vowel characters were *inconsistent* and had two alternative pronunciations. The *inconsistent-conditioned* pronunciation occurred only following a particular consonant character, whilst the *inconsistent-unconditioned* pronunciation occurred following any of the other consonant characters. Each of these consistency levels (consistent vs. inconsistent-conditioned vs. inconsistent-unconditioned) had a low and a high frequency form, occurring in four and eight items respectively. Adults were

not taught about individual character-sound mappings, only about whole-word pronunciations.

This paradigm allowed exploration of four main questions. First, can learners extract sub-word regularities from exposure to whole-word orthographic forms (and their corresponding pronunciations) without explicit instruction, and use this to support generalization? Second, is knowledge extracted about spelling-sound mappings context-sensitive? Third, are learning and generalization influenced by the frequency and predictability of spelling-sound mappings present in the language environment? Finally, do learners develop whole-item orthographic representations and are these also sensitive to spelling-sound frequency and predictability? These questions will now be addressed in turn.

Extraction of character-sound mappings

Examination of performance during training demonstrated that participants were able to learn to read the novel words to at least 80% accuracy, although the ease with which they did so varied. Some participants read 80% of items correctly in the second block of training whilst others required up to seven blocks. Following training, generalization performance was high with participants reading on average 73% of the novel test items correctly. Error analyses indicated that by the end of training, the majority of incorrect responses to both trained and generalization items resulted from using the alternative pronunciation for inconsistent vowel characters. These findings show very clearly that learners extracted the sounds of individual characters following exposure to whole-word forms, and that they were able to use this knowledge to read

novel items successfully. This provides support for the learning process embodied in the triangle model of reading aloud in which implicit knowledge of spelling-sound mappings develops through corrected whole-word pronunciation attempts.

One interpretation of these findings might be that as participants were highly literate adults, they were simply mapping the artificial orthography onto their knowledge of English letter-sound mappings. This seems unlikely, especially for vowels: the vowel sounds in the novel orthography can be represented in multiple ways in written English. For example, the phoneme /i/ can be spelled, E, EE, EA, IE, EI, EY, and /u/ can be spelled, U, O, OO, UE, OU, UI, EW, etc. Therefore, vowel character-phoneme relationships could not be directly mapped to English.

Context-sensitivity

The spelling-sound knowledge that participants extracted was indeed context-sensitive. During training, participants learned to use the conditioned or unconditioned pronunciation of inconsistent vowel characters according to the consonant context. This was also the case, although to a somewhat lesser extent, in generalization. These findings demonstrate that context-sensitive spelling-sound knowledge develops through exposure to the orthographic forms of whole words and their pronunciations without explicit instruction. Again this supports predictions made by the triangle model in which the implicit knowledge of spelling-sounds that develops through exposure is not restricted to single letter-sounds but takes account of surrounding letters and their

pronunciations. This context-sensitivity helps the model to read words such as BOOK, TOOK, and LOOK differently from words such as FOOD, SPOON, and ZOOM, for example.

One concern might be that the context-sensitive rules that governed inconsistent vowel pronunciations in the artificial orthography were deterministic. Perhaps this made them easier to abstract than the probabilistic conditional relationships that exist in English (Treiman et al., 1995)? However, the less than perfect generalization speaks against this. In addition, post-experiment debriefing demonstrated that none of the participants were explicitly aware of the consonant-vowel pronunciation rules. In sum, learning the novel orthography was achievable but not trivial, and most participants were not able to identify the context-sensitive rules that they had nevertheless extracted.

Frequency and consistency effects

Turning to the third question, the frequency and predictability of character-sound mappings had clear influences on learning. Items containing high frequency vowels outperformed those containing low frequency vowels throughout training, and this effect was maintained in the block at which 80% accuracy was achieved. In addition, this effect was only significant in inconsistent and not consistent items. Although the current study used a type frequency manipulation, these results can still be considered to mirror those from English orthography, in which words with inconsistent spelling-sound mappings are read faster when they are high frequency than when they are low frequency by both adults (Jared, 2002; Paap & Noel, 1991; Seidenberg et al., 1984) and children (Weekes et al., 2006).

Within items containing low frequency vowels, learning was also influenced by consistency in a predictable way. Consistent items were easier to learn than inconsistent items, supporting findings from research on English orthography (Jared et al., 1990; Weekes et al., 2006). Additionally, items containing the conditioned inconsistent vowel type were easier to learn than items containing the unconditioned inconsistent vowel type. Inconsistent-unconditioned items had the least predictable spelling-sound mappings because the vowel pronunciation occurred following a variety of consonant onsets. Inconsistent-conditioned items had more predictable spelling-sound mappings as the vowel pronunciation followed a specific consonant character. This finding is therefore in line with Kessler and Treiman's (2001) demonstration that the predictability of vowel pronunciations in English words is increased when the consonant context is taken into account. Overall these findings demonstrate that the more predictable a spelling-sound pattern, the more easily it is learned.

Somewhat unexpected results were obtained for items containing high frequency vowels. At the beginning of training, performance was higher on inconsistent-conditioned items than on both consistent and inconsistent-unconditioned items. It seems likely that the repeated onset-vowel combination that characterised items containing high frequency inconsistent-conditioned items (e.g., in stimulus set 1, 8 out of 36 items contained the onset /vi/) was salient to participants and that this inflated performance on such items. By the middle-to-end of training, performance on consistent and inconsistent-conditioned items was equivalent and performance on inconsistent-unconditioned items remaining poorest. This suggests that, with extended

exposure, participants became sensitive to the fact that consistent items contained a vowel character with an unvarying pronunciation and that this was beneficial to learning.

Consistency also had predictable effects on generalization. Accuracy was higher when reading novel test items containing vowel characters that were consistent during training than novel items containing vowel characters that were inconsistent during training. However, unlike in training, generalization performance was unaffected by vowel frequency. In addition, there was no difference between performance on items that should have taken the inconsistent-conditioned pronunciation and those that should have taken the inconsistent-unconditioned pronunciation. This suggests that while learning of the training set was context and frequency sensitive, this sensitivity was less well retained for use in generalization.

This might be taken as evidence to support the DRC model of reading aloud which uses context-insensitive spelling-sound rules for reading novel words. However, this model would predict that only the most frequent pronunciation of each inconsistent character should be used in generalization. This was not the case. When generalization data were rescored such that only the most frequent pronunciation of inconsistent vowel characters was accepted as correct, performance levels were poorer than when only the appropriate conditioned or unconditioned pronunciation was accepted as correct. This demonstrates that pronunciation of generalization items was context-sensitive, albeit to a lesser extent than pronunciation of trained items.

Item-level representations

The final question was whether whole-item orthographic representations of the training set developed. Recognition of the orthographic forms of trained items was above chance but there was a high number of false positive responses to untrained items. In addition, response latencies were extremely long (in the order of seconds rather than milliseconds). This suggests that participants did not automatically recognise trained items; instead they decoded them and based their decision on a combination of orthographic and phonological familiarity. This is very different to the automatic recognition skilled readers display in lexical decision, as evidenced by the fact that RTs in such tasks are generally around 700ms (Balota et al., 2004; Lacruz & Folk, 2004; Stone et al., 1997). Participants are also likely to have inaccurately decoded some items and this may have led them to make incorrect judgements as to whether the item was trained or untrained. Overall, it can be concluded that participants did not develop strong item-specific representations of the orthographic forms of the training set.

In lexical decision using English orthography, accuracy and latencies are sensitive to item frequency and consistency (Lacruz & Folk, 2004; Stone et al., 1997). In contrast, old-new decisions in the artificial orthography were unaffected by the frequency and consistency characteristics of trained items. Coupled with the long RTs, this suggests that although old-new decision provides a gross measure of item recognition, it is a laborious task and is less sensitive to lexical variables than typical lexical decision.

Conclusions

Experiment 2 provided the first evidence that adults can learn to read an artificial orthography in a relatively short experimental paradigm. Although naming and recognition of trained items was fairly laboured, learners did extract context-sensitive spelling-sound knowledge and use this to support generalization. In addition, effects of frequency and consistency emerged from exposure to the orthography, influencing learning and, to a somewhat lesser extent, generalization in ways that largely mirrored those seen in English orthography. These results provide the first conclusive evidence that knowledge of context-sensitive sub-word spelling-sound mappings can be extracted through exposure to the orthographic form of whole words and their pronunciations, as predicted by the triangle model of reading aloud. Although the artificial orthography paradigm cannot be said to mimic all the processes involved in learning to read, it provides a high degree of experimental control and generates data that simulate typical reading processes. This makes it an ideal tool to address further issues in word reading research.

Several points deserve further investigation. Although generalization was quite good, participants did not use context-sensitive pronunciations of inconsistent vowel characters to the same extent as in training. In addition, discrimination of trained from untrained items was laboured and there were high numbers of false positive responses. Old-new decisions were also insensitive to the frequency and consistency characteristics of trained items and sensitivity to these variables was also less pronounced in generalization than during training. It seems likely that extended training would improve

both generalization and discrimination. In Experiment 3, a larger group of participants are trained to 70% proficiency on one day and to 90% proficiency on the following day, with the prediction that generalization and discrimination will improve and show increased sensitivity to frequency and consistency characteristics extracted from the training set.

CHAPTER 4: THE INFLUENCE OF PROFICIENCY ON DISCRIMINATION AND GENERALIZATION

This experiment explored whether increased experience in reading the artificial orthography would improve discrimination of trained from untrained items and generalization to novel items, relative to Experiment 2. It also explored whether this extended training would induce/strengthen frequency and consistency effects in discrimination and generalization. Sixteen adults learned to read the artificial orthography to 70% and 90% accuracy on consecutive days. Discrimination and generalization were improved on Day 2. Day 2 also saw the emergence of consistency effects in discrimination. However, even on Day 1, discrimination was enhanced relative to that seen in Experiment 2. Furthermore, frequency influenced discrimination on Day 1, and generalization was sensitive to frequency and consistency on both days. The major improvements on Experiment 2 therefore seem to be increased power (more participants) and clearer instructions, rather than increased experience with the artificial orthography.

Experiment 2 introduced a novel method for investigating the factors influencing orthographic learning. Adults learned to read novel words written in novel characters which varied in the frequency and predictability of their character-sound mappings. This *artificial orthography learning paradigm* gave precise control over the input statistics to

which learners were exposed. This avoided several of the problems that have plagued previous research using English orthography; including reliance on lexical databases, dichotomization of non-categorical variables, and the influence of non-controlled factors such as age-of-acquisition.

Using this artificial orthography paradigm it was shown that learners could extract individual character-sounds through exposure to whole-word forms, and that they were sensitive to conditional relationships between consonant onsets and vowel character pronunciations. Items containing vowels with higher frequency and more consistent character-sound mappings were read more accurately throughout training. Frequency and consistency also interacted in their effects: frequency only influenced items containing inconsistent vowels and consistency only influenced items containing low frequency vowels. This mirrors findings from English orthography showing that both adults and children are more accurate and faster when reading words which are high in frequency and/or have consistent spelling-sound mappings (Jared, 2002; Weekes et al., 2006). At the end of training, participants could discriminate trained from untrained items at above chance levels and were able to generalize and read novel test items successfully. Like learning, generalization also showed evidence of context-sensitivity; learners pronounced vowels in accordance with their consonant onset more often than with an alternative pronunciation. This again reflects data from English orthography showing that when reading nonwords, both adults (Treiman et al., 2003) and children (Treiman et al., 2006) are aware of how vowels are pronounced in different consonant contexts.

Issues arising from Experiment 2

Although the artificial orthography paradigm provided a high degree of experimental control and generated data that simulate typical reading processes, several issues demand further investigation, two of which are the focus of the current study. First, discrimination of trained from untrained items (as assessed by old-new decision) differed in several ways from discrimination of words and nonwords in lexical decision tasks in natural languages. Second, participants were less sensitive to the frequency and consistency of character-sound mappings when reading generalization items than they were in training. The implications of these findings will now be discussed in more detail.

Discrimination

Although discrimination of trained from untrained items was above chance, false positive error rates were high, and responses to both trained and untrained items were extremely laboured, with decision latencies in the order of seconds. This is very different to typical lexical decision in which RTs are usually around 700ms (Balota et al., 2004; Lacruz & Folk, 2004; Stone et al., 1997). Furthermore, discrimination was not affected by the frequency and consistency characteristics of trained items. In contrast, in typical lexical decision, both frequency (Balota & Chumbley, 1984; Balota et al., 2004; Cortese & Khanna, 2007) and consistency (Lacruz & Folk, 2004; Stone et al., 1997) have been shown to influence performance. Additionally, Lacruz and Folk (2004) reported a

consistency by frequency interaction in lexical decision: consistency effects were present for low but not high frequency words.

Overall, these findings suggest that old-new decision differed from lexical decision in that participants did not automatically recognise the orthographic forms of the items they had learned to read. Instead they had to decode them and use a combination of orthographic and phonological familiarity to consciously evaluate whether they had learned the item. In some ways this difference is not surprising. Participants had only 30-45 minutes exposure to the novel orthography whereas adults have had a lifetime of experience with words. However, it would clearly be of interest to examine experience dependent changes in discrimination.

Unfortunately, very little research has focused on such issues. Castles and colleagues found that lexical decision error rates and RTs were greater in children than adults (Castles, Davis, Cavalot, & Forster, 2007; Davis, Castles, & Iakovidis, 1998). In particular, Castles, Davis, and Letcher (1999) reported that performance improves between the ages of 8 and 12 years, but has still not reached adult levels. Regarding the emergence of frequency and consistency effects in lexical decision, research in fact suggests that sensitivity to these variables decreases with increasing experience. For example, Waters, Seidenberg, and Bruck (1984) reported that frequency and regularity effects in lexical decision are more pronounced in children than adults, and in poor readers relative to good readers. Similarly, Unsworth and Pexman (2003) found that consistency effects in lexical decision were more pronounced in less skilled adult readers than in more skilled adults. Finally, Waters et al. reported that while regularity effects

were restricted to low frequency words in adults, they were present for both low and high frequency words in younger poor readers. A point to note about Waters et al.'s investigations is that regularity effects were only present when words with irregular spelling-sound correspondences included some "strange" words, defined as words with unusual spellings, such as GAUGE, AISLE or SEIZE. They argue that this reduces the extent to which lexicality judgements can be made on the basis of orthographic familiarity, which increases phonological processing.

Overall, these findings suggest that discrimination error rates and RTs should decrease with increasing experience. However, predictions perhaps remain open regarding the emergence of frequency and consistency effects in old-new decision in an artificial orthography. On the one hand, research on English orthography suggests that such effects should diminish with experience, but on the other, participants have undoubtedly had far less experience with trained items than even children or poor readers have had with real words. Therefore, it still might be the case that further experience is required for frequency and consistency effects to influence discrimination of trained items in an artificial orthography.

Generalization

Experiment 2 found that generalization accuracy was lower than trained item reading accuracy. This difference between end of training and generalization performance was largely driven by items containing inconsistent vowels. Specifically, 76% of inconsistent items were read correctly at the end of training and only 63% were read correctly in generalization, whereas for consistent items, end of training accuracy

was 96% and generalization accuracy was 93%. This suggests that while learners had extracted the sounds of the individual characters, their sensitivity to the conditional relationships between consonant onsets and vowel character pronunciations was somewhat item-specific and did not wholly transfer from training to generalization. In addition, in generalization, items containing conditioned and unconditioned inconsistent vowels were read equally poorly relative to items containing consistent vowels, whereas during training participants found the conditioned pronunciation of inconsistent vowel characters easier to learn. Furthermore, generalization accuracy was insensitive to vowel frequency. The fact that sensitivity to the frequency and predictability of character-sound mappings did not entirely transfer from training to generalization again suggests that there was some element of item-specific learning during training.

As with discrimination, it is likely that the ability to extract context-sensitive character-sound mappings and use them to generalize increases with experience. This is supported by developmental evidence. Even fairly young children are sensitive to consistency when reading words (Weekes et al., 2006) but their ability to use such knowledge in novel word reading tasks increases over time. Treiman et al. (2006) examined 7-, 9-, and 11-year-olds' reading of irregular-consistent nonwords and regular-consistent nonwords. Regular-consistent nonwords contain an orthographic body that is pronounced according to GPCs in the majority of English words, e.g., POOM (ROOM, ZOOM, LOOM) or YASH (DASH, CASH, MASH), whereas in irregular-consistent nonwords the orthographic body is not pronounced according to GPCs in most English words, e.g., POOK (GPC = /u/ vs. BOOK, TOOK, LOOK) or CHIND (GPC = /i/ vs. FIND, MIND, KIND).

Treiman et al. found that although all age groups were more likely to use an irregular pronunciation for irregular-consistent nonwords than for regular-consistent nonwords, the frequency of irregular pronunciations for the former increased between the ages of 7 and 11. This demonstrates that the use of surrounding context to predict vowel pronunciation increases with experience. Thus, further training should increase overall accuracy in generalization, largely through increased accuracy on items containing inconsistent vowels. It may also strengthen the influence of vowel frequency and consistency.

Experiment 3

The aim of this experiment was to investigate whether extended training improved discrimination and generalization and strengthened the effects of vowel frequency and consistency in these tasks. The experiment used the same procedure and orthography as Experiment 2. However, participants received two training sessions on two consecutive days. On Day 1, participants learned to read the novel words to at least 70% accuracy, and on Day 2 they received further training until they achieved at least 90% accuracy. Tests of discrimination and generalization followed training on each day. It was predicted that discrimination and generalization performance would be enhanced on Day 2 relative to Day 1, and also that frequency and consistency effects would be stronger in both tasks on Day 2. Given that trained item reading accuracy would be almost at ceiling by the end of Day 2, the influence of frequency and consistency on trained items was expected to be reduced on Day 2 relative to Day 1.

Method

Participants

Sixteen adults (6 males, 10 females) took part. They were all university students and their mean age was 21.5 years, $SD = 2.5$. All participants had English as a first language and had normal or corrected to normal hearing and vision. None had participated in Experiment 2.

Materials and Procedure

All participants completed the same procedure which took place at the same time on two consecutive days. On each day, training and testing were completed in 30-45 minutes. On Day 1, the experiment took much the same form as Experiment 2 and consisted of an orthography exposure and learning phase, followed by old-new decision and generalization tasks. However, in the learning phase, participants were required to read at least 70% of items correctly by the end of training, in contrast to the 80% criterion level set in Experiment 2. The same procedure was used on Day 2, except that no exposure phase was required and participants were trained to 90% accuracy during the learning phase. Additionally, 24 different untrained items (12 for each stimulus set) were designed for use in generalization and as distractors in old-new decision on Day 2 (see Appendix 4.A). Note that old-new decision also included different trained items on Day 2 than on Day 1.

A further alteration was that before starting the old-new decision task on each day, participants were instructed that they would probably find the task very difficult because items they had and hadn't learned would look extremely similar. It was hoped

that this instruction would enhance performance by explicitly making participants aware that they had to decide whether a whole item was familiar, as opposed to individual characters.

Results

Training

Individual differences in learning. Figure 4.1 shows performance during training on Day 1 and Day 2. If a participant required less than 6 blocks to achieve criterion (Day 1 - 70%, Day 2 - 90%), accuracy in remaining blocks was assumed to be identical to that in the final block of training completed. The mean number of blocks to achieve 70% accuracy on Day 1 was 3.63 ($SD = 1.46$), with a maximum of 6 and a minimum of 2. There was a highly significant negative correlation between accuracy in Block 1 and the number of blocks required to achieve 70% accuracy, $r = -.72, p < .01$.

There was no drop in performance between the last block of training on Day 1 and the first block of training on Day 2, $t_p(15) = 1.23, ns$. The mean number of blocks required to achieve 90% accuracy on Day 2 was 2.88 ($SD = 1.15$), again with a maximum of 6 and a minimum of 2. The number of blocks required to achieve 90% accuracy did not correlate with accuracy in Block 1 of Day 2, accuracy in Block 1 of Day 1 or the number of blocks required to achieve 70% accuracy on Day 1. However, accuracy in Block 1 of Day 2 did correlate with the number of blocks required to achieve 70% accuracy on Day 1, $r = -.54, p < .05$.

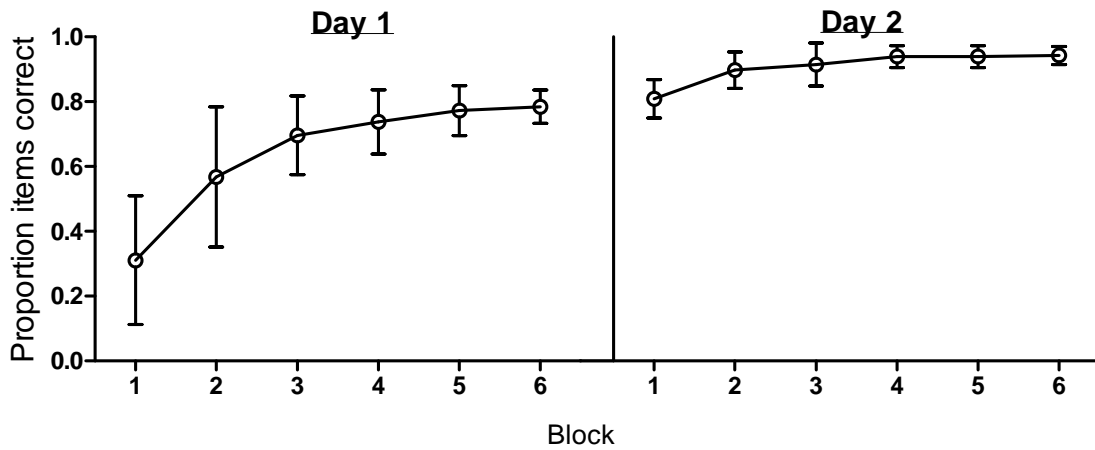


Figure 4.1. Proportion of items read correctly (\pm SD) during training on Day1 and Day 2.

Frequency and consistency effects at the end of training on Day 1 and Day 2. Figure 4.2 shows that clear frequency and consistency effects and an interaction between these two variables were observed at the end of training on Day 1. It can also be seen that, as predicted, these effects were less pronounced on the second day, as by this point performance was almost at ceiling ($\geq 90\%$).

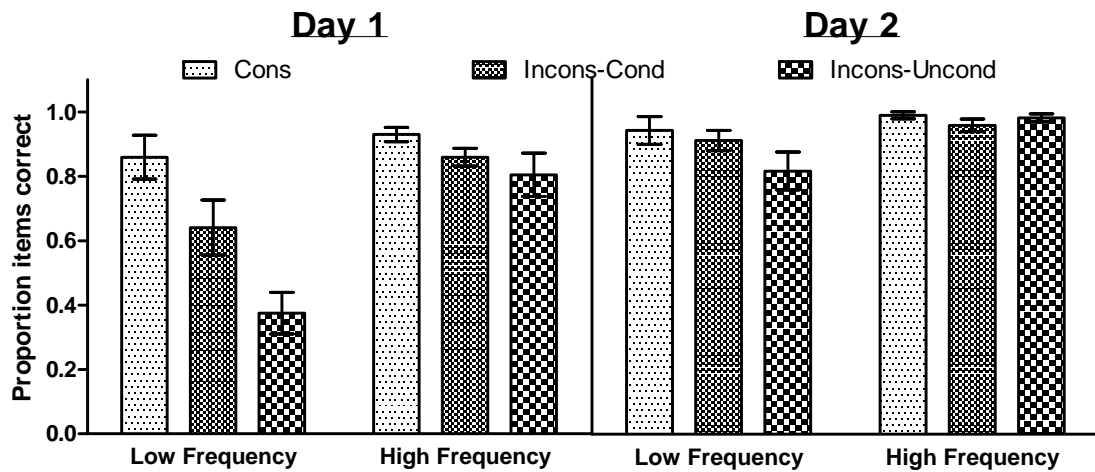


Figure 4.2. Proportion of items read correctly (\pm SE) at the end of training on Day1 and Day 2, as a function of vowel frequency and consistency.

The effects of vowel frequency and consistency at the end of training were assessed by conducting an analysis of variance on the proportion of items read correctly in the block at which 70% accuracy was achieved on Day 1 vs. the block at which 90% accuracy was achieved on Day 2. On both days this constituted different blocks for different participants. In the subjects analysis, frequency, consistency, and day were treated as within subject factors, whereas in the items analysis, frequency and consistency were treated as between items factors and day as a within item factor.

Reflecting the requirement for participants to achieve 70% accuracy at the end of Day 1 and 90% accuracy at the end of Day 2, accuracy was higher on Day 2 than Day 1, $F_p(1, 15) = 102.16, p < .001, \eta_p^2 = .87, F_i(1, 66) = 89.83, p < .001, \eta_p^2 = .58$. Performance was better for items containing high frequency vowels than for items containing low frequency vowels $F_p(1, 15) = 34.19, p < .001, \eta_p^2 = .60, F_i(1, 66) = 38.06, p < .001, \eta_p^2 = .37$. Accuracy was higher for consistent than inconsistent-conditioned items, which in turn outperformed inconsistent-unconditioned items, $F_p(2, 30) = 12.47, p < .001, \eta_p^2 = .45, F_i(2, 66) = 16.50, p < .001, \eta_p^2 = .33$. There was also a significant interaction between frequency and consistency, $F_p(2, 30) = 4.24, p < .05, \eta_p^2 = .22, F_i(2, 66) = 7.12, p < .01, \eta_p^2 = .18$. Pairwise comparisons indicated that the facilitative effect of frequency was only significant within inconsistent items, and that the consistency effect was only significant within items containing low frequency vowels.

The effects of frequency and training day interacted, $F_p(1, 30) = 6.59, p < .05, \eta_p^2 = .31, F_i(1, 66) = 15.73, p < .001, \eta_p^2 = .19$. Pairwise comparisons and examination of effect sizes indicated that the effect of frequency was significant on both days but that it was more pronounced on Day 1. This was largely because the increase in accuracy on Day 2 was more pronounced for items containing low frequency vowels. The effect of consistency also interacted with training day, $F_p(2, 30) = 6.57, p < .01, \eta_p^2 = .31, F_i(2, 66) = 12.62, p < .001, \eta_p^2 = .28$. On Day 1, the effect of consistency was as reported for the main effect, whereas on Day 2, accuracy was higher for consistent items than inconsistent-unconditioned items, but performance on inconsistent-conditioned items fell in between. In addition, the improvement between Day 1 and 2 was most marked in inconsistent-unconditioned items, followed by inconsistent-conditioned items, and was least marked in consistent items. The three-way interaction between frequency, consistency and day was significant by-items, $F_i(2, 66) = 3.31, p < .05, \eta_p^2 = .09$, but not by-subjects, $F_p(2, 30) = 1.57, ns$. Pairwise comparisons across items demonstrated that this interaction stemmed from the fact that on Day 1, the effect of frequency was significant in both conditioned and unconditioned inconsistent items, whereas on Day 2 it was only significant in inconsistent-unconditioned items.

Old-new decision

Trials with RTs more than two standard deviations from a participant's mean were excluded from the analysis (Day 1: 3.6%, Day 2: 3.9%). On Day 1, one participant had no correct responses to items containing low frequency inconsistent-conditioned

vowels, and another had no correct responses to items containing high frequency inconsistent-unconditioned vowels. On Day 2, one participant had no correct responses to items containing high frequency inconsistent-conditioned vowels and another had no correct responses to items containing high frequency inconsistent-unconditioned vowels. These missing data points were replaced with the mean RT for that item type.

Table 4.1 shows old-new decision accuracy and RTs to trained and untrained items on Days 1 and 2. It also shows results from one sample t-tests demonstrating that accuracy was above chance for both trained and untrained items on both days.

Table 4.1.

Old-new decision accuracy and RTs to trained and untrained items and one-sample t-tests to confirm accuracy above chance.

Item type	Response measure	<i>M</i>	<i>SD</i>	<i>t</i> (15)	<i>p</i>
<u>Day 1</u>					
Trained	Accuracy	.82	.09	14.48	< .001
	RT (ms)	3775	989		
Untrained	Accuracy	.63	.12	4.32	.001
	RT (ms)	4743	1369		
<u>Day 2</u>					
Trained	Accuracy	.88	.11	21.65	< .001
	RT (ms)	2944	544		
Untrained	Accuracy	.83	.09	14.35	<.001
	RT (ms)	3831	754		

On Day 1 accuracy was significantly better for trained than untrained items, $t_p(15) = 4.64, p < .001$. RTs were also faster to trained than untrained items, $t_p(15) = 5.08, p < .001$. On Day 2, accuracy was equivalent for trained and untrained items, $t_p(15) = 1.42, p = .18$. However, RTs remained faster to trained than untrained items, $t_p(15) = 6.53, p < .001$.

Figures 4.3 (accuracy) and 4.4 (RTs) summarise the effects of frequency and consistency on old-new decision performance on Days 1 and 2. Accuracy was slightly higher and RTs were faster on Day 2 than on Day 1. Accuracy does not appear to have been affected by the frequency or consistency characteristics of trained items on either day. However, RTs were faster to items containing high frequency vowels, particularly on Day 1. RTs also seem to have been faster to consistent than inconsistent items, particularly when items contained low frequency vowels.

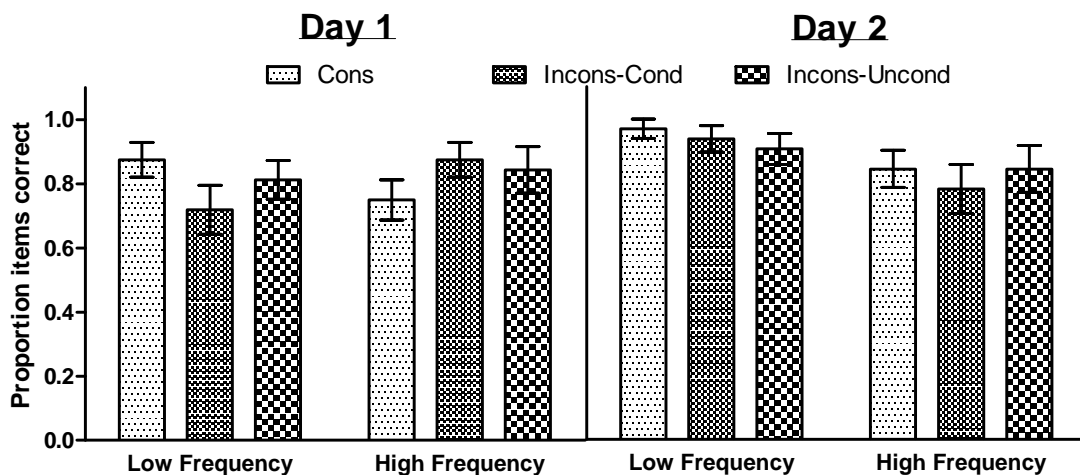


Figure 4.3. Old-new decision accuracy ($\pm SE$) as a function of frequency and consistency

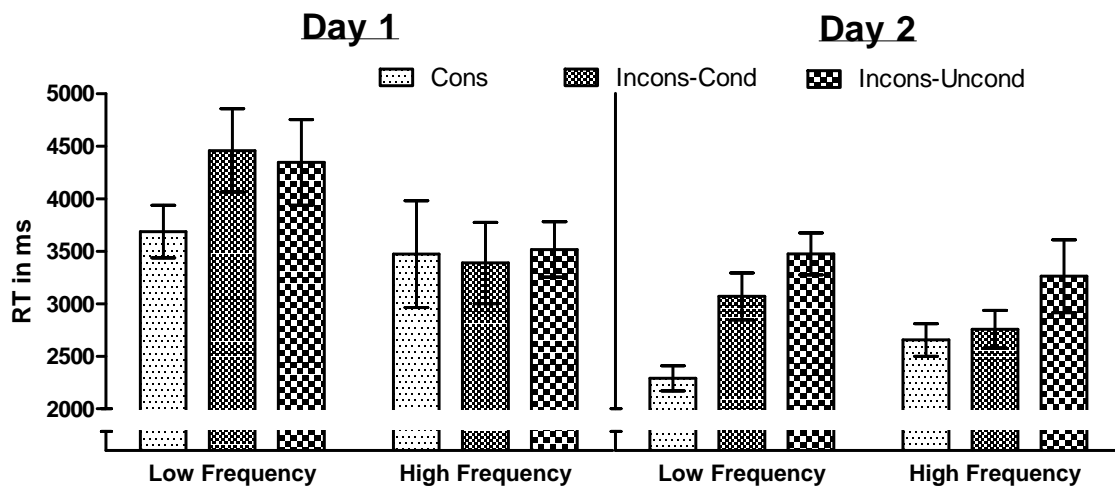


Figure 4.4. Old-new decision RTs in ms (\pm SE) as a function of frequency and consistency

Two analyses of variance were conducted to assess the effects of vowel frequency and consistency on old-new decision performance on Day 1 and Day 2; one using proportion of trained items correct as the dependent variable and one using mean RTs to correct trained items. In the subjects analysis, frequency, consistency, and day were treated as within subject factors, whereas in the items analysis, these factors were between items. Note that day of testing was treated as a between items factor in the items analysis because different items were used for testing on Days 1 and 2.

Accuracy was higher on Day 2 than on Day 1, although this effect was only marginally significant by-subjects, $F_p(1, 15) = 4.35$, $p = .06$, $\eta_p^2 = .23$, and was not significant by-items, $F_i(1, 36) = 2.26$, $p = .14$, $\eta_p^2 = .06$. The enhanced performance on Day 2 was reflected in the RT analysis, $F_p(1, 15) = 19.50$, $p < .001$, $\eta_p^2 = .57$, $F_i(1, 36) = 23.79$, $p < .001$, $\eta_p^2 = .40$. There were no other main effects or interactions in the accuracy analysis (see Appendix 4.B).

In the RT analysis there was a main effect of frequency, $F_p(1, 15) = 4.39, p = .05, \eta_p^2 = .23, F_i(1, 36) = 6.76, p = .01, \eta_p^2 = .16$, that interacted with day of testing, $F_p(1, 15) = 4.25, p = .06, \eta_p^2 = .22, F_i(1, 36) = 4.64, p < .05, \eta_p^2 = .11$. Pairwise comparisons confirmed that RTs were faster to items containing high frequency vowels than to items containing low frequency vowels on Day 1 but not Day 2, and that the reduction in RTs between Days 1 and 2 was more pronounced for items containing low frequency vowels. RTs were faster to consistent than inconsistent-unconditioned items, with RTs to inconsistent-conditioned items falling in between, $F_p(2, 30) = 5.47, p < .01, \eta_p^2 = .27, F_i(2, 36) = 5.53, p < .01, \eta_p^2 = .24$. This consistency effect did not interact with day of testing, $F_p(2, 30) = 1.71, ns, F_i(2, 36) = 1.33, ns$, or with frequency, $F_p(2, 30) = 1.59, ns, F_i(2, 36) = 1.52, ns$, and there was no three-way interaction between these variables, $F_s < 1$. However, pairwise comparisons revealed that consistency effects were only present on Day 2 and in items containing low frequency vowels.

Generalization

Figure 4.5 summarises the effects of vowel frequency and consistency on generalization accuracy on Days 1 and 2. Examination of this figure suggests that frequency and consistency had similar effects on Days 1 and 2, with a frequency by consistency interaction being evident on both days. Within items containing low frequency vowels, accuracy was higher when vowels were consistent than when they were inconsistent-unconditioned. This effect was less pronounced for items containing high frequency vowels on Day 1, and was negligible in such items on Day 2. On both

days the frequency effect was greatest in items containing inconsistent-unconditioned vowels.

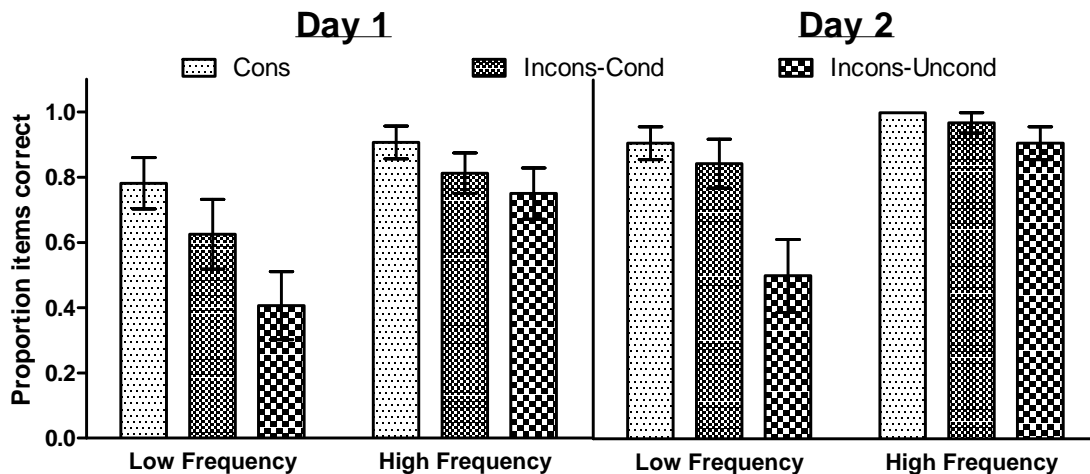


Figure 4.5. Generalization accuracy ($\pm SE$) as a function of frequency and consistency.

An analysis of variance was conducted to examine the effects of vowel frequency and consistency during training on the proportion of generalization items read correctly on Days 1 and 2. As in old-new decision, frequency, consistency, and day of testing were within subject factors in the subjects analysis, and between items factors in the items analysis. Performance was better on Day 2 than Day 1, $F_p(1, 15) = 59.75$, $p < .001$, $\eta_p^2 = .80$, $F_i(1, 36) = 12.37$, $p = .001$, $\eta_p^2 = .26$. Accuracy was higher for items containing high frequency vowels than for items containing low frequency vowels, $F_p(1, 15) = 15.16$, $p = .001$, $\eta_p^2 = .50$, $F_i(1, 36) = 28.49$, $p < .001$, $\eta_p^2 = .44$. This frequency effect did not interact with day of testing, $F_s < 1$. Performance was equivalent for items containing consistent and inconsistent-conditioned vowels and both outperformed items containing inconsistent-unconditioned vowels, $F_p(2, 30) = 10.44$, $p < .001$, $\eta_p^2 = .41$, $F_i(2, 36) = 14.36$,

$p < .001$, $\eta_p^2 = .44$. Again consistency did not interact with day of testing, $F_s < 1$. There was also an interaction between frequency and consistency, $F_p(2, 30) = 3.43$, $p = .05$, $\eta_p^2 = .19$, $F_i(2, 36) = 4.19$, $p < .05$, $\eta_p^2 = .19$, which did not enter into a three-way interaction with day of testing, $F_s < 1$. Pairwise comparisons indicated that the effect of frequency was only significant in items containing inconsistent-unconditioned vowels, and that the effect of consistency was only significant in items containing low frequency vowels.

Discussion

The aim of this experiment was to examine the influence of extended training on old-new decision and generalization. In Experiment 2, discrimination of trained from untrained items was laboured and subject to false positive responding. Discrimination accuracy and reaction times were also unaffected by the frequency and consistency characteristics of trained items. Generalization accuracy was fairly high, but the pattern of performance suggested that participants were less sensitive to the frequency and predictability of character-sound mappings than they were in training. Children's performance on lexical decision (Castles et al., 2007; Castles et al., 1999; Davis et al., 1998; Unsworth & Pexman, 2003; Waters et al., 1984) and novel word reading tasks (Treiman et al., 2006) suggests that extended experience with the novel orthography might improve discrimination and generalization, and induce frequency and consistency effects in these tasks. The current experiment sought to investigate this possibility.

Sixteen adults learned to read the artificial orthography in two training sessions on two consecutive days. They achieved at least 70% accuracy on the training set on Day

1 and at least 90% accuracy on Day2. On both days, testing measures included reading accuracy at the end of training, discrimination of trained from untrained items, and generalization. Items presented in the discrimination and generalization tasks were different on Days 1 and 2. The results of these outcome measures will now be discussed in turn.

Trained item reading

At the end of training on Day 1, performance was better for items containing high frequency vowels than for those containing low frequency vowels. Accuracy was also higher for consistent items followed by inconsistent-conditioned items, and was lowest for inconsistent-unconditioned items. These effects were qualified by a frequency by consistency interaction: frequency only influenced items containing inconsistent vowels, and consistency only influenced items containing low frequency vowels. These findings replicate those of Experiment 2 when participants were required to achieve 80% accuracy at the end of training. Somewhat surprisingly, the frequency by consistency interaction persisted at the end of training on Day 2 although, at this level of proficiency, consistent and inconsistent-conditioned items were read with equal accuracy and frequency only influenced inconsistent-unconditioned items. These persisting significant effects were unexpected given that performance was practically at ceiling. These findings demonstrate that frequency and consistency exert strong, reliable, and predictable influences on reading accuracy in the artificial orthography. This further supports its use in investigating the factors which influence word reading.

Discrimination

The first point to note is that on Day 1, accuracy was higher for trained and untrained items than it was in Experiment 2. Of particular importance is the fact that, unlike in Experiment 2, accuracy for untrained items was above chance on Day 1. Explicitly informing participants that trained and untrained items would be very similar and that the task would be difficult seems to have increased their understanding of the task and improved performance. As predicted, old-new decision error rates and RTs were reduced on Day 2 relative to Day 1. Thus, increased experience with the artificial orthography did indeed increase participants' ability to recognise individual items.

Discrimination accuracy remained unaffected by the frequency and consistency characteristics of trained items. However, RTs were influenced by frequency and consistency in a predictable way. On Day 1, RTs were faster to items containing high frequency vowels than to those containing low frequency vowels. The fact that this effect emerged at 70% proficiency suggests that Experiment 2 failed to obtain frequency effects in old-new decision due to a lack of understanding of the task, and perhaps a lack of power due to smaller participant numbers, rather than lack of exposure to the training set. Frequency influenced RTs when participants were less proficient (Day 1) but not when they were more proficient (Day 2). This is in accordance with studies on lexical decision in English demonstrating that frequency effects reduce with increasing reading skill/experience (Waters et al., 1984).

RTs were fastest to consistent items and slowest to inconsistent-unconditioned items. Thus, consistency exerted a similar effect on old-new decision performance to

that observed in typical lexical decision (Lacruz & Folk, 2004; Stone et al., 1997). Intermediate RTs were observed for inconsistent-conditioned items, which had somewhat more predictable character-sound mappings. This supports research demonstrating that consistency is a graded variable (Jared et al., 1990). Consistency effects did not in fact emerge until Day 2, which is perhaps counterintuitive given that in lexical decision tasks using English words, consistency effects decrease with increasing experience/proficiency (Unsworth & Pexman, 2003; Waters et al., 1984). However, as suggested in the introduction, participants in the current experiment received far less exposure to the training items than even beginning readers have to real words. This finding does not therefore go against existing research on English orthography. It does however suggest that sensitivity to frequency and consistency may develop with different amounts of exposure, with frequency effects emerging and dissipating earlier than consistency effects. This could be further investigated by conducting longitudinal investigations of the emergence of frequency and consistency effects in lexical decision in children.

Although the interaction between consistency and frequency in old-new decision was not significant, the consistency effects seen on Day 2 were in fact only robust in items containing low frequency vowels. This again mirrors results from English orthography demonstrating that with increasing reading skill, consistency effects are only observed in low frequency words (Lacruz & Folk, 2004; Waters et al., 1984). Altogether, the predictable influences of frequency and consistency on old-new decision performance observed in the current experiment support the idea that many of the

factors that influence orthographic processing in natural languages also influence processing of trained items in the artificial orthography.

It is important to consider what drives the experience-dependent changes seen in old-new and typical lexical decision. Castles and colleagues (2007) suggested that orthographic representations are underspecified in young readers and become more specified with increasing exposure/proficiency. This idea has been investigated by combining lexical decision with *masked priming*. In a masked priming paradigm, targets and distractors are preceded by a very short priming stimulus (~50ms), followed by a mask. The relationship between the prime and target can be systematically manipulated. Forster and Davis (1984) established that decision latencies to word targets are decreased by prior presentation of the same word, an *identity prime*, relative to a different word, a *control prime*. Castles et al. (1999) extended this to show that when primes are minimally different from the target, e.g., *kest* or *ebst* to prime *BEST*, adults do not show a priming effect if targets have many neighbours (words that can be created by changing just one letter), e.g., *BEST* - *NEST*, *REST*, *BENT*, *BUST*. In contrast, children show facilitation for minimally different primes for words from both sparse and dense neighbourhoods (Castles et al., 2007). This supports the idea that children's orthographic representations are more broadly tuned than those of adults. It is possible that a similar tuning process occurred with increasing experience with the artificial orthography.

Although trained item processing in old-new decision was subject to many of the same influences as real words in typical lexical decision, old-new decisions still took

between 2 and 3 seconds. This is far slower than latencies seen in typical lexical decision. This is hardly surprising given that participants only received a maximum of 90 minutes exposure to the novel orthography, whereas even a 7-year-old child will have had around 2 or 3 years daily experience with letters. A further factor to consider is that trained and untrained items were all consonant-vowel-consonant monosyllables, constructed from 16 consonant and 4 vowel characters. Even when restricted to monosyllables, lexical decision tasks use English words of between four and six letters with a wider variety of consonant and vowel combinations. This will have increased the difficulty of old-new decision because items in the artificial orthography were more similar to each other and more confusable than real words in natural languages.

However, it is still likely that trained item and real word representations differ in some ways. Participants did not automatically recognise the orthographic forms of trained items; instead they seemed to decode them and consciously evaluate their phonological and orthographic familiarity. This is different to the automatic recognition typically displayed in lexical decision in natural languages. Aside from amount of experience, one potentially important difference between real words and trained items is that the real words have meaning. This is addressed in Experiment 4, which explores the impact of endowing trained items with meaning on learning and subsequent processing. The similarity of trained item and real word processing is also further explored in the next chapter by examining whether old-new decisions to trained items are facilitated by masked identity priming. This is motivated by research demonstrating that lexical decisions to words (but not nonwords) are faster when preceded by an

identical masked priming stimulus, than when preceded by a masked control stimulus (Bowers, 2003; Davis et al., 2008; Forster et al., 2003; Forster, 1998; Forster & Davis, 1984). If trained items are sensitive to masked priming this will support the idea that they are processed in a similar way to words in natural languages.

Generalization

Developmental research suggests that sensitivity to the consonant context when pronouncing vowels in novel words increases between the ages of 6 and 11 (Treiman et al., 2006). In Experiment 2, learners displayed less context-sensitivity in generalization than in training. In the current experiment, it was predicted that further training would improve generalization. It was expected that this improvement would largely arise from increased accuracy when reading generalization items containing inconsistent vowels. A further prediction was that extended training would increase the effects of vowel frequency and consistency on generalization.

The first of these predictions was borne out. Participants read an average of 71% of items correctly on Day 1 and 85% on Day 2. However, contrary to expectations, this improvement was not restricted to items with context-dependent vowel pronunciations. Accuracy increased from 64% to 80% for items containing inconsistent vowels, and from 84% to 95% for items containing consistent vowels. This suggests that further training increased both knowledge of individual character sounds, and sensitivity to the conditional relationships between consonant onsets and vowel character pronunciations.

Furthermore, sensitivity to the frequency and consistency of vowel pronunciations was equivalent on Days 1 and 2 and mirrored that seen in training. Specifically, frequency facilitated performance on items containing inconsistent vowels, and consistency facilitated performance on items containing low frequency vowels. In addition, items containing the more predictable inconsistent-conditioned vowels were read more accurately than those containing the less predictable inconsistent-unconditioned vowels. These results demonstrate that participants were sensitive to the frequency and predictability of character-sound mappings in the training set even on Day 1. It is possible that such effects did not emerge in Experiment 2 due to a lack of power from smaller participant numbers, rather than due to a lack of exposure to the training set.

It should be noted here that, as in Experiment 2, the majority of participants were not aware of the rules that governed inconsistent vowel pronunciations. On Day 1, post-experiment debriefing indicated that none of the participants had explicitly deduced the rules. At the end of Day 2, one participant had noticed the onset-vowel relationship that characterised items containing high frequency inconsistent-conditioned vowels. The others still had no explicit knowledge of inconsistent vowel pronunciation rules, despite achieving at least 90% accuracy during training and an average of 84% in generalization. This implicit context and frequency sensitive character-sound knowledge can be considered to be compatible with the way in which spelling-sound knowledge is represented in the triangle model of reading aloud (Harm & Seidenberg, 2004; Plaut et al., 1996).

Conclusions

In the current experiment, frequency and consistency interacted in their effects on trained item reading, replicating the results of Experiment 2, and mirroring studies using English orthography. The strong, robust, and predictable effects obtained using the artificial orthography paradigm support its use for further investigating the factors influencing orthographic processing. As predicted, old-new decision performance improved following further exposure to the novel orthography. Although RTs were still long, they were influenced by the frequency and consistency characteristics of trained items in a way which largely mirrored effects reported in typical lexical decision. Generalization also improved with extended training and, on both days, the pattern of performance indicated that learners had extracted context-sensitive character-sound mappings and used them to support generalization.

Overall, the results of the current experiment strongly support the idea that implicit knowledge of context-sensitive spelling-sound mappings emerges through exposure to whole-word orthographic forms and their pronunciations, as suggested by the triangle model of reading aloud (Harm & Seidenberg, 2004; Plaut et al., 1996). Discrimination and generalization ability increased in conjunction, and displayed similar sensitivity to the frequency and predictability of character-sound mappings. This is consistent with the idea that the orthography-phonology mappings used for generalization and the item-specific knowledge required to recognise trained items were supported by the same representations, again as embodied in the triangle model.

One important difference between real words and trained items in the current experiment is that trained items had no meaning. This is addressed in Experiment 4 by the introduction of a pre-exposure phase in which participants learn the associations between the phonological form of trained items and a novel object referent. The impact of pre-exposure on learning, discrimination, and generalization is then assessed. The similarity of trained and real word processing is also directly examined by incorporating masked priming into old-new decision.

CHAPTER 5: THE ROLE OF SEMANTICS IN LEARNING TO READ

In the triangle model, semantics supports reading aloud, particularly for low frequency inconsistent words. However, behavioural support for this position is not unequivocal. The current experiment utilised the artificial orthography paradigm to investigate the role of semantics in learning to read. 32 participants learned to read the artificial orthography to at least 70% accuracy. 16 participants were pre-exposed to the association between item sounds and novel objects (semantics) and 16 participants were pre-exposed to item sounds (lexical phonology). Relative to a no pre-exposure baseline (Day 1, Experiment 3), semantic pre-exposure increased reading accuracy, particularly for items containing low frequency inconsistent vowels. Semantic pre-exposure also enhanced discrimination but hindered generalization. However, lexical phonology pre-exposure had equivalent effects. This suggests that knowledge of word meanings does not facilitate orthographic learning, beyond the contribution provided by item-specific phonological representations. A qualification to this is that stronger semantic representations may be necessary for effects to outweigh those of phonological familiarity.

The role of semantics in orthographic processing was addressed in Experiment 1 in which adults learned to read novel words written in English orthography. Although results provided some evidence for an influence of semantics on reading accuracy,

effects were not reflected in reading or discrimination latencies. Furthermore, because novel words were written in English orthography, reading proficiency was very high from the beginning of training. Thus, Experiment 1 failed to address the question of how semantics influences *learning* to read. It also meant that the experiment was not immune to many of the problems which plague natural language research, such as measuring and controlling for spelling-sound consistency. In addition, although research suggests that the role of semantics in reading aloud varies according to word frequency and spelling-sound consistency, existing data provide very little insight into how we develop sensitivity to these lexical statistics.

Experiments 2 and 3 introduced an artificial orthography. This novel method was used to investigate how the frequency and predictability of spelling-sound mappings influence a) learning to read, b) the development of whole-item representations, and c) generalization. Learners extracted knowledge of context-sensitive spelling-sound patterns through exposure to whole words and used this knowledge successfully in generalization. Learning, discrimination, and generalization were influenced by the frequency and consistency of spelling-sound mappings in a way that largely mirrored that reported in studies of English orthographic processing (Jared, 2002; Jared et al., 1990; Lacruz & Folk, 2004; Treiman et al., 2003). The pattern of results was compatible with the idea that the same representations support our ability to read and recognise specific items and to generalize. This is consistent with the process of learning to read embodied in the triangle model (Harm & Seidenberg, 2004; Plaut et al., 1996) in which

implicit knowledge of spelling-sound patterns is developed through exposure to whole-word orthographic forms and their pronunciations.

Experiments 2 and 3 established that the frequency and consistency of spelling-sound mappings influence learning to read in the artificial orthography in a predictable way. However, there was some evidence to suggest that trained items were not processed in the same way as words in natural languages. In particular, discrimination of trained from untrained items was extremely laboured indicating that strong whole-item orthographic representations had not developed. It is possible that the absence of meaning in the language contributed to this. The current experiment thus returned to the question of how semantics influences orthographic learning.

Reasons for postulating semantic involvement in orthographic processing are first recapped, framed in terms of how models of reading aloud accommodate semantics. A critique of studies demonstrating semantic effects on orthographic processing is then presented. This is relatively brief as this literature was discussed in greater depth in Chapter 2. A discussion then follows of the difficulty of disentangling the influences of semantic and phonological familiarity and how previous experiments have attempted to address this issue. Finally, the method used in the current experiment is outlined and justified.

How might semantics support orthographic processing?

In Experiments 2 and 3, adults found novel words harder to learn when they contained character-sound mappings that were low in frequency and somewhat unpredictable. This mirrors research on English orthography demonstrating that words

with atypical spelling-sound mappings are read aloud with more difficulty by both adults (Balota et al., 2004; Cortese & Simpson, 2000; Jared et al., 1990; Parkin, 1982; Waters & Seidenberg, 1985) and children (Seymour et al., 2003; Weekes et al., 2006). This effect is also more pronounced for low frequency words (Jared, 2002; Paap & Noel, 1991; Seidenberg et al., 1984; Weekes et al., 2006). Together these findings provide clear evidence that our reading system is sensitive to the commonality and typicality of spelling-sound patterns and that this develops through exposure to whole-word forms.

Although relative difficulties with low frequency inconsistent words are seen in experimental tasks, a skilled reader of English is nonetheless able to read such words to a high degree of proficiency. As described in Chapters 1 and 2, the triangle model of reading aloud (Harm & Seidenberg, 2004; Plaut et al., 1996) suggests that knowledge of word meanings supports this ability, because semantics can directly aid the computation of phonology from orthography. In contrast, in the Dual Route Cascaded model (DRC - Coltheart et al., 2001) irregular word reading is possible because whole-word orthography-phonology mappings are stored by the lexical route. Although Coltheart et al. proposed that the lexical route is subdivided into a semantic and a non-semantic pathway, the semantic lexical route is considered to be too slow to be of use in skilled adult reading. Furthermore, because the lexical route is also used to read regular words once they have been experienced enough times, the DRC makes no specific predictions about the importance of semantics for reading regular vs. irregular words.

Both the DRC and triangle model suggest that word meanings can also impact on the process of visual word recognition. The DRC recognises known words because the

lexical route stores whole-word orthographic representations. There are both semantic and non-semantic pathways in this lexical route, allowing for word meanings to impact on the recognition process. In the triangle model there are reciprocal connections between semantic and orthographic representations and thus word meanings again have the opportunity to influence visual word recognition (Harm & Seidenberg, 2004; Plaut, 1997; Plaut et al., 1996). Unlike in reading aloud, no explicit predictions have been made as to whether semantic variables interact with frequency or consistency in lexical decision in the triangle model.

Evidence for semantic involvement in orthographic processing

As discussed in Chapter 2, research findings provide some support for the triangle model's proposed role for semantics in orthographic processing, but this evidence is not unequivocal. Neuropsychological patients with impairments in semantic memory often have difficulties with reading aloud, which are most pronounced when reading low frequency inconsistent words (McKay et al., 2007; Patterson et al., 2006; Woollams et al., 2007). However, some patients with semantic dementia do not show deficits in reading aloud (Blazely et al., 2005; Gerhand, 2001), questioning the direct relationship between semantics and inconsistent word reading embodied in the triangle model. In addition, semantic dementia is a degenerative condition which makes it difficult to rule out the possibility that reading deficits are caused by the extension of damage to other language regions, such as those that represent orthography.

Strain et al. (1995) and Frost et al. (2005) have shown that imageability, a semantic variable, influences the accuracy and speed with which adults read low

frequency exception words. Large scale regression studies have also found semantic variables such as imageability, meaningfulness, the number of other words with which a word is associated, and semantic richness to predict naming and lexical decision latencies (Balota et al., 2004; Cortese & Khanna, 2007; Pexman, Hargreaves, Siakaluk, Bodner, & Pope, 2008). Note that in lexical decision no interaction between semantics and spelling-sound consistency has been reported. Corroborating evidence is provided by developmental research. For example, Ricketts et al. (2007) found that expressive vocabulary skills, one index of individual differences in word-level semantic knowledge, predicted reading accuracy of exception words, but not regular words or nonwords, in a sample of 9-year-old children. Thus, semantic influences on reading aloud that are specific to inconsistent words seem to begin fairly early in development.

A difficulty for natural language experiments, such as those described above, is the reliance on lexical databases to extract ratings of frequency, consistency/regularity, and semantic variables such as imageability. For example, Strain et al. (1995) took imageability ratings from the Medical Research Council Psycholinguistic Database (Coltheart, 1981) which were derived from a merging of several other databases (Gilhooley & Logie, 1980; Paivio, Yuille, & Madigan, 1968; Toggia & Battig, 1978). However, ratings were unavailable for 11 of the words used in the study and these had to be rated by members of the authors' research institution. Thus, imageability ratings came from different sources for different words. Furthermore, different studies have used different imageability ratings, for example, Cortese and Khanna (2007) used imageability ratings developed by Cortese and Fugett (2004). It is likely that individual

differences exist in the degree to which words are considered imageable, but such differences are not typically examined. The likelihood of variation between individuals means that ratings may differ according to the sample from which they are obtained. This in turn means that results of different studies are not directly comparable.

Strain et al. (1995) state that "... Finding words with the appropriate characteristics of frequency and regularity that also had imageability ratings proved to be a severe limitation..." (p. 1141). This was in part a result of the factorial design they employed in which words were defined as high or low frequency, regular or exception, and high or low imageability. In fact, this design is problematic in other ways because variables that are graded in nature (and in their effects) were artificially categorised. As discussed in Chapter 3, this has several consequences, such as differences in category boundaries between experiments, correlations between category variables which may cause analyses to overestimate differences between groups, and restricted item lists which are liable to idiosyncratic effects (MacCallum et al., 2002). Thus, reliance on factorial designs may have yielded unreliable effects of semantics on orthographic processing.

A rather different concern is how to operationalize semantics. Many studies have used imageability as an index of word meaning. However, Balota et al. (2004) found that *Nelson's set size*; the number of associated words produced by two or more subjects in free association, and *meaningfulness*; the number and strength of other words that come to mind, were also significant predictors of lexical decision performance, although they did not influence naming. Similarly, Pexman et al. (2008) found that the number of

words appearing in similar lexical contexts had a greater effect on lexical decision than the number of semantic features, despite the fact that both variables are considered to be measures of semantic richness. These findings illustrate that semantics is a complicated construct and suggest that studies which only consider imageability (or indeed any other single semantic variable) may not capture all the variance in orthographic processing associated with word meaning. When all these problems are taken into consideration, it becomes clear that the evidence for a role for semantics in orthographic processing is by no means definitive and warrants further investigation.

Semantics or item-level phonology?

A further problem that plagues research on semantics is that the correlation between familiarity with word meanings and familiarity with word sounds is often ignored. This issue was first highlighted by Monaghan and Ellis's work (Ellis & Monaghan, 2002; Monaghan & Ellis, 2002) showing that the influence of imageability on reading aloud (Strain et al., 1995;2002) disappeared once age-of-acquisition was taken into account. Note that this conclusion is also supported by Cortese and Khanna's (2007) re-analysis of Balota et al.'s (2004) large scale regression study. Although there is some controversy as to whether AoA effects are largely phonological or semantic in nature (Bonin et al., 2004; Cortese & Khanna, 2007; Morrison & Gibbons, 2006; Morrison et al., 2002), it remains likely that early acquired words will have stronger phonological representations (Brown & Watson, 1987; Morrison & Ellis, 1995; Morrison et al., 1992). Thus, if AoA accounts for imageability effects, the purported effect of semantics on

reading aloud may in fact be driven by familiarity with word sounds rather than familiarity with word meanings.

The issue of semantic vs. phonological familiarity was considered in a novel word learning study by McKay et al. (2008, Experiment 2), as discussed in Chapter 2. Pre-exposure to definitions of novel words (semantics) increased subsequent reading accuracy and speed relative to pre-exposure to the sounds of novel words (lexical phonology). This benefit was specific to novel words that had been assigned inconsistent spelling-sound mappings. McKay et al. also found that semantic pre-exposure improved performance in an old-new decision task in which participants discriminated trained from untrained novel words. This effect was not specific to inconsistent items, mirroring research on English orthography where an interaction between consistency and semantics in lexical decision has not been reported.

McKay et al.'s (2008) study seems to provide clear evidence that semantic familiarity supports orthographic processing over and above any benefit obtained through phonological familiarity. However, there are several issues that warrant further consideration. First, a baseline no pre-exposure condition was not included therefore it is possible that phonological familiarity did benefit subsequent orthographic processing, albeit to a lesser extent than semantic familiarity. Second, the semantic manipulation was within subject; participants learned the definitions of semantic items and this task was interspersed with blocks in which they listened to and repeated the sounds of lexical phonology items. An efficient strategy would therefore have been to concentrate on the demanding task of learning the definitions in the semantic condition and to pay

minimal attention during the listen-repeat task in the lexical phonology condition. Thus, McKay et al. may have underestimated the benefit of phonological familiarity for word reading and recognition and, in turn, overestimated the importance of semantic familiarity. Finally, as in Experiment 1 in this thesis, novel words were written in English orthography and reading accuracy was therefore extremely high from the beginning. In fact McKay et al. explicitly state that "... Accuracy on consistent novel words was at ceiling from the first trial..." (p. 1500). As a result, this study failed to address the role of semantics in *learning* to read.

McKay et al. argued that trained novel words were treated similarly to words in natural languages on the basis that naming latencies for trained but not untrained items were decreased by masked identity priming. However, as discussed in Chapter 1, the specificity of masked identity priming effects to words and not nonwords is only robust in lexical decision and not naming tasks (Bowers, 2003; Forster et al., 2003). Thus, it is debatable to what extent McKay et al.'s (2008) observations inform us about the similarity of newly learned and real word processing. This would be clearer if masked identity priming effects specific to trained (and not untrained) items could be shown in old-new decision.

Experiment 4

Outline of method

This experiment utilised the control provided by the artificial orthography paradigm to investigate the role of semantic vs. phonological familiarity in orthographic

learning. It also examined the extent to which trained items in the novel orthography were treated similarly to words in natural languages. Thirty-two participants learned to read the artificial orthography to at least 70% accuracy, using the procedure described in Experiment 3 - Day 1. Prior to this, half the participants were pre-exposed to the sounds of the items (*lexical phonology*), and half were pre-exposed to their sounds plus a novel object referent (*semantics*).

As in Experiments 2 and 3, outcome measures included reading accuracy during training, old-new decision, and generalization. In old-new decision, items were preceded by a masked identity or control prime to assess whether masked priming effects were present for, and specific to, trained and not untrained items in the artificial orthography. Performance following semantic vs. lexical phonology pre-exposure was compared to the no pre-exposure baseline provided by Day 1 of Experiment 3.

Justification

This method avoided many of the problems identified in previous research. First, the no pre-exposure baseline enabled direct examination of any benefits obtained from phonological familiarity alone. Second, semantic vs. lexical phonology pre-exposure was manipulated between subjects, meaning that learning semantic associations could not interfere with attending to lexical phonology item sounds. Attention during pre-exposure was also ensured by intermittently asking participants whether a particular item had been repeated in the previous five trials. Third, using the artificial orthography paradigm meant that semantic effects could be examined as participants were *learning* to read. In contrast, most previous research has focused on the influence of semantics in

skilled reading; even developmental research has failed to directly investigate semantic effects on learning. Furthermore, previous word learning studies have used novel words written in English orthography and thus also assessed fairly skilled reading. Experiments 2 and 3 established that strong, reliable, and predictable effects of spelling-sound frequency and consistency emerge when learning to read an artificial orthography. The artificial orthography paradigm therefore also avoids problems associated with extracting variables from lexical databases, categorizing graded variables, and the influence of non-controlled factors such as AoA.

An issue that deserves some consideration is the use of novel object referents to represent semantics. This decision was motivated by several factors. It was considered important that participants learn new meanings for training items, rather than associating them with an existing real word item, for example, by teaching them that /buv/ was a giraffe. Using real word meanings would have increased the likelihood of participants assimilating trained items to English words. Furthermore, novel definitions were felt to be problematic because existing research suggests that word meanings vary on numerous dimensions, such as imageability, meaningfulness, the number of associated words, and the number of semantic features (Balota et al., 2004; Pexman et al., 2008). Thus, it would have been difficult to develop 36 novel definitions that were equated (or varied systematically) along these dimensions. Finally, the literature on infant word learning suggests that knowing the association between an object and a label represents an early hallmark of understanding word meanings. For example, while young infants can discriminate minimal phonetic contrasts (Kuhl, Williams, Lacerda,

Stevens, & Lindblom, 1992; Werker & Tees, 1984), this ability is compromised in older infants when they must also attend to learning word meanings (Stager & Werker, 1997; Werker, Fennell, Corcoran, & Stager, 2002). In such studies learning *word meanings* involves learning *word-object pairings*. Similarly, to investigate whether infants link the process of statistically segmenting words from fluent speech to that of mapping meanings to words, Estes, Evans, Alibali, and Saffran (2007) tested whether newly segmented words could be associated with objects. Werker, Cohen, Lloyd, Casasola, and Stager (1998) suggest that although the ability to map labels to objects is not equivalent to that of linking words to past experiences/ideas as is characteristic of adult knowledge of word meanings, it does represent "... a significant advance in the child's progression toward word learning..." (p. 1307). Thus, it was felt that teaching novel object-label associations would capture a core feature of learning word meanings.

Predictions

Pre-exposure to semantics (novel objects) was expected to increase reading accuracy for items containing low frequency inconsistent vowels and to improve old-new decision performance for all items, relative to the no pre-exposure baseline provided by Day 1 of Experiment 3. The existing literature suggests that pre-exposure to lexical phonology might also improve reading accuracy and discrimination, relative to no pre-exposure. However, it was predicted that facilitation in the lexical phonology condition would be less pronounced than in the semantic condition. Trained but not untrained items were expected to benefit from masked identity priming in old-new

decision in both pre-exposure conditions. Generalization performance was not expected to be influenced by pre-exposure.

Method

Participants

Thirty two adults (15 males, 17 females) participated. Their mean age was 20.13 years, $SD = 3.17$. All participants had English as a first language and had normal, or corrected to normal, hearing and vision. None had participated in Experiment 2 or 3.

Materials and procedure

All participants learned to read the artificial orthography to at least 70% accuracy, using the same procedure as described in Experiment 3 - Day 1. Prior to orthography training, half the participants were pre-exposed to *lexical phonology* and half to *semantics*. The experimental procedure is summarised in Figure 5.1 and described in more detail below.

Lexical phonology pre-exposure. Participants listened to and repeated each of the 36 training items five times and were instructed to try and remember how they sounded. To ensure that they remained on task, every seven to ten trials they were asked whether the previous item had also occurred in the five trials that had come before it. Presentation was randomised across all 180 trials.

Semantic pre-exposure. Word meanings were represented by photographs of novel objects (see Appendix 5.A for some examples). These were taken from a photo objects database and were judged to be unknown by 20 adults who did not take part in

the current experiment. Each object was randomly assigned to one of the 36 training items. Four different random assignments were used and a quarter of participants experienced each in order to minimise any effects of a particular item being easier to associate with a particular novel object. Participants listened to and repeated each of the 36 training items while viewing the associated novel objects. They were instructed to try and remember the names of the objects. To ensure that they remained on task, every seven to ten trials they were asked whether the previous item had also occurred in the five trials that had come before it. Each item and associated object appeared four times and presentation was randomised across all 144 trials. A fifth and final exposure trial was used to assess semantic learning. Participants were presented with four of the novel objects and were asked to decide which one matched a spoken item. Each item appeared once with its novel object referent and presentation order was randomised. The spatial location of referents and distractors was pseudorandomised and was the same for each participant. No feedback was given.

Retention of semantic knowledge was tested following orthography training and old-new decision. Participants were presented with four of the novel objects and the *written* form of one of the training items. They were required to select the object that matched the written item. As well as semantic knowledge retention, this also assessed participants' ability to decode the item well enough to match it with its object referent. Each item appeared once with its novel object referent and presentation order was randomised. The spatial location of referents and distractors was pseudorandomised and was the same for each participant. No feedback was given.

Old-new decision. Following orthography training (and the retention of semantic knowledge test in the semantic condition) participants completed the old-new decision task. This took the same form as described in Experiment 3 - Day 1, but was modified to incorporate masked priming. Sets of 24 identity and 24 control primes were developed. Identity primes were the same as the targets, whereas control primes contained the same first character as the targets but different second and third characters (see Appendix 5.B for phonological forms of control primes). In typical lexical decision, primes are presented in lower case and targets in uppercase. To approximate this, both types of prime were a third smaller than targets.

A mask (###) appeared on screen for 450ms, followed by a blank screen for 50ms, the priming stimulus for 50ms, and finally the target which remained on screen until a response was made. Instructions were the same as for Experiment 3, with participants being told that they would find the task difficult because trained and untrained items would look very similar. They were not told about the presence of the primes, but were informed that “before each word, some ### marks will flash on the screen. Just ignore them and respond to the word.”

Generalization took exactly the same form as in Experiment 3 - Day 1. Pre-exposure, training, and testing were completed in a single session lasting between 45 and 60 minutes.

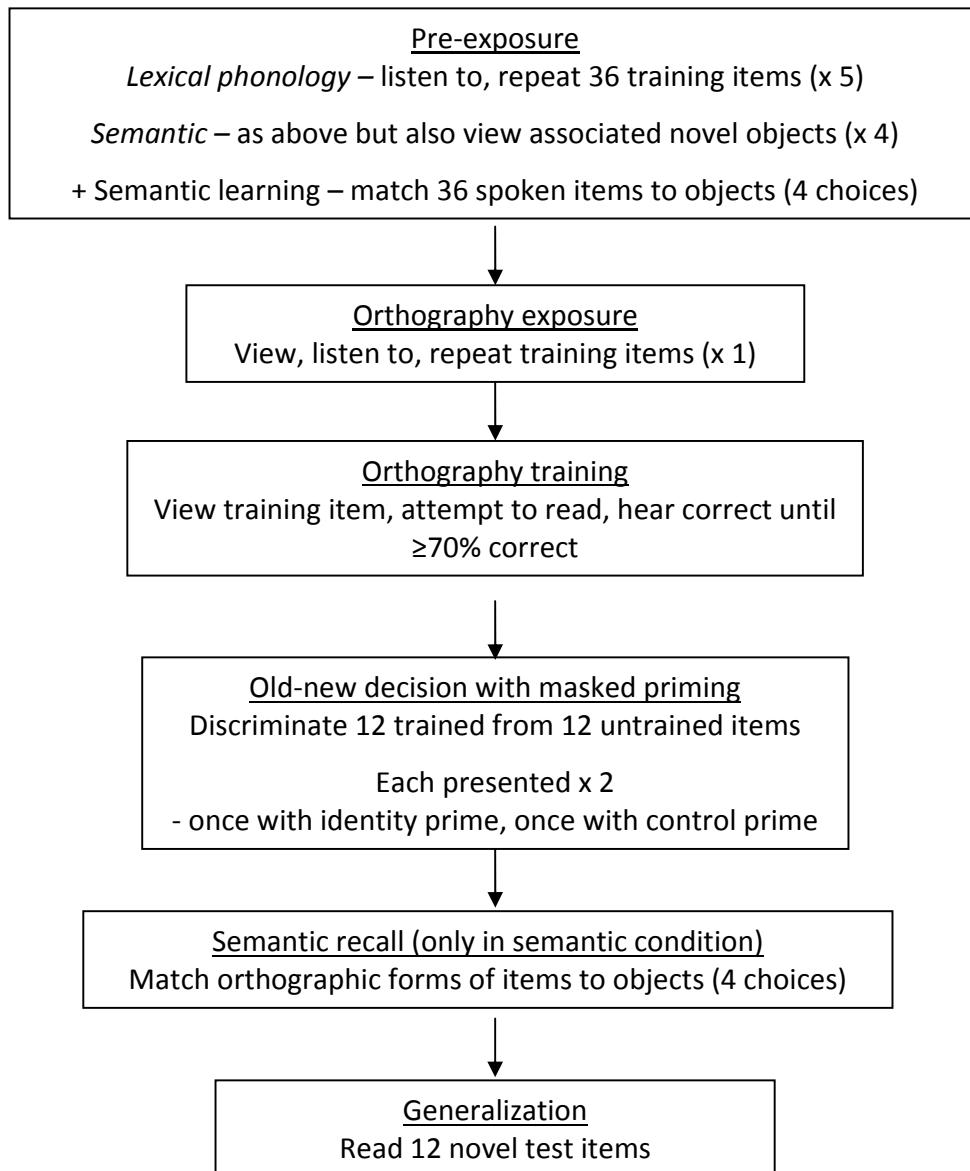


Figure 5.1. Summary of procedure for Experiment 4.

Results

Semantic learning

It was first important to ensure that participants in the semantic condition had learned the associations between the training items and the novel objects. The mean

proportion of items correct in the spoken semantic learning task which was administered directly after pre-exposure was .88 ($SD = .12$). In the written semantic learning task which was completed following orthography training, the mean proportion of items correct was .83 ($SD = .15$). One samples t-tests confirmed that in both cases this was significantly above chance performance levels of .25 (spoken: $t_p(15) = 21.74, p < .001, t_i(71) = 40.59, p < .001$, written: $t_p(15) = 15.76, p < .001, t_i(71) = 30.33, p < .001$) demonstrating that the item-object associations had indeed been learned and retained.

Training

Effect of pre-exposure on overall performance during training. The maximum number of blocks required to achieve criterion was 5 in the lexical phonology and semantic conditions and 6 in the no pre-exposure condition (Experiment 3 - Day 1). Figure 5.2 summarises performance in each condition over the course of training. Where participants required less than 6 blocks to achieve criterion, accuracy in remaining blocks was assumed to be identical to that in the final block of training completed. The mean number of blocks to achieve the criterion of 70% accuracy (SD in parentheses) was 2.63 (1.31) in the lexical phonology condition and 2.75 (1.24) in the semantic condition. As reported in Experiment 3 (Day 1), the mean number of blocks to achieve 70% accuracy in the no pre-exposure condition was 3.63 (1.14). There was a highly significant negative correlation between accuracy in Block 1 of training and the number of blocks required to achieve 70% accuracy, in both the lexical phonology, $r = -.76, p = .001$, and semantic conditions, $r = -.77, p < .001$. As reported in Experiment 3 this

was also the case in the no pre-exposure condition, $r = -.72$, $p < .01$. These observations suggested that the pattern of learning was broadly similar across pre-exposure groups.

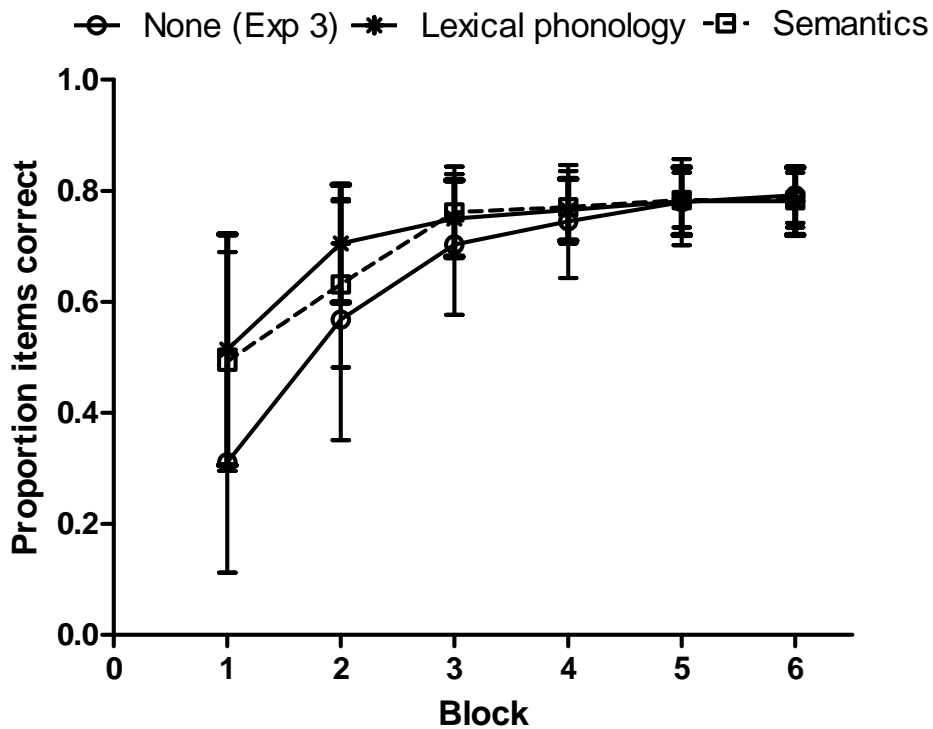


Figure 5.2. Accuracy during training (\pm SD) as a function of pre-exposure.

The overall effect of pre-exposure was assessed by conducting a repeated measures analysis of variance on the proportion of items read correctly in each block of training. In the subjects analysis, pre-exposure (none vs. lexical phonology vs. semantics) was treated as a between subjects factor and block (Block 1 - Block 6) as a within subject factor, whereas for the items analysis, both block and pre-exposure were treated as within item factors. Reading accuracy improved over the course of training in all conditions, as shown by the main effect of block, $F_p(1.77, 79.53) = 94.05$, $p < .001$, $\eta_p^2 = .68$, $F_i(1.93, 136.81) = 255.97$, $p < .001$, $\eta_p^2 = .78$. Reading accuracy was facilitated by pre-

exposure to semantics or lexical phonology relative to no pre-exposure, $F_p(2, 45) = 3.17$, $p = .05$, $\eta_p^2 = .12$, $F_i(2, 142) = 7.56$, $p = .001$, $\eta_p^2 = .10$. The effects of block and pre-exposure interacted, $F_p(10, 225) = 3.54$, $p < .001$, $\eta_p^2 = .14$, $F_i(10, 710) = 14.29$, $p < .001$, $\eta_p^2 = .17$. Pairwise comparisons indicated that pre-exposure provided significant facilitation in only the first and second blocks of training. Also, note that although performance in the no pre-exposure condition continued to improve in each subsequent block of training, it reached asymptote by Block 3 in the lexical phonology condition and by Block 4 in the semantic condition.

Pre-exposure, frequency, and consistency effects at the beginning of training.

Figure 5.3 shows the effect of pre-exposure when reading items containing vowels of differing frequency and consistency in the first block of training. Pre-exposure to either lexical phonology or semantics seems to have boosted performance across item types, relative to the no pre-exposure condition provided by Day 1 of Experiment 3. The effects of pre-exposure (none vs. lexical phonology vs. semantics), frequency (high frequency vs. low frequency), and consistency (consistent vs. inconsistent-conditioned vs. inconsistent-unconditioned) were assessed by conducting an analysis of variance on the proportion of items read correctly in block 1 of training. In the subjects analysis, frequency and consistency were treated as within subject variables and pre-exposure as a between subjects variable, whereas in the items analysis, frequency and consistency were between items variables and pre-exposure was within item. These designs were also used to analyse performance at the end of training, reported in the next section.

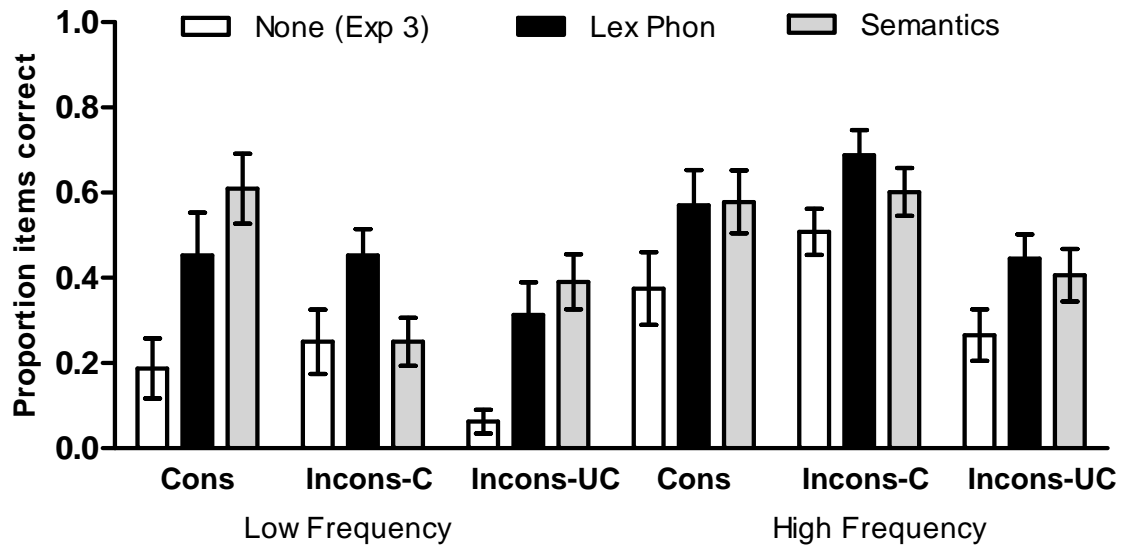


Figure 5.3. Accuracy at the beginning of training ($\pm SE$) as a function of pre-exposure, frequency, and consistency.

Accuracy was higher in the lexical phonology and semantic conditions than in the no pre-exposure condition, $F_p(2, 45) = 5.91, p < .01, \eta_p^2 = .21, F_i(2, 132) = 34.80, p < .001, \eta_p^2 = .35$. Performance was better for items containing high frequency vowels than for those containing low frequency vowels, $F_p(1, 45) = 51.32, p < .001, \eta_p^2 = .53, F_i(1, 66) = 29.90, p < .001, \eta_p^2 = .31$, and for consistent and inconsistent-conditioned items than inconsistent-unconditioned items, $F_p(2, 90) = 11.98, p < .001, \eta_p^2 = .21, F_i(1, 66) = 10.71, p < .001, \eta_p^2 = .25$. The main effects of frequency and consistency interacted, $F_p(2, 90) = 6.99, p < .01, \eta_p^2 = .14, F_i(1, 66) = 3.98, p < .05, \eta_p^2 = .11$. Pairwise comparisons and examination of effect sizes indicated that the facilitative effect of frequency was significant across item types, but that it was weakest in consistent items, followed by

inconsistent-unconditioned items and strongest in inconsistent-conditioned items. In addition, within items containing high frequency vowels, accuracy was higher for consistent and inconsistent-conditioned items than for inconsistent-unconditioned items, as reported for the main effect, but within items containing low frequency vowels, accuracy was higher for consistent than inconsistent-unconditioned items, with inconsistent-conditioned items falling in between.

The effect of pre-exposure interacted with consistency, $F_p(4, 90) = 3.04, p < .05, \eta_p^2 = .12, F_i(4, 132) = 4.50, p < .01, \eta_p^2 = .12$. Pairwise comparisons demonstrated that within consistent and inconsistent-unconditioned items, the effect of pre-exposure was as reported for the main effect. However, within inconsistent-conditioned items, accuracy was higher in the lexical phonology condition than in the no pre-exposure condition with accuracy in the semantic condition falling in between. In addition, in the lexical phonology and no pre-exposure conditions, the effect of consistency was as reported for the main effect, but in the semantic condition, accuracy was higher for consistent items than for both types of inconsistent items, which did not differ from each other. The effect of pre-exposure did not interact with frequency, $F_p(2, 45) = 1.74, p = .18, \eta_p^2 = .07, F_i(2, 132) = 1.68, p = .19, \eta_p^2 = .03$, and there was no three-way interaction between frequency, consistency, and pre-exposure, $F_p(4, 90) = 1.88, p = .12, \eta_p^2 = .08, F_i(4, 132) = 1.77, p = .14, \eta_p^2 = .05$.

Pre-exposure, frequency, and consistency effects at the end of training. Figure 5.4 shows the effect of pre-exposure on reading accuracy in the final block of training. The

facilitative effect of pre-exposure now appears to be restricted to items containing low frequency inconsistent-unconditioned vowels. The effect of pre-exposure at the end of training was assessed with an analysis of variance on the proportion of items read correctly in the block at which 70% accuracy was achieved. It should be remembered that this constituted different blocks for different participants.

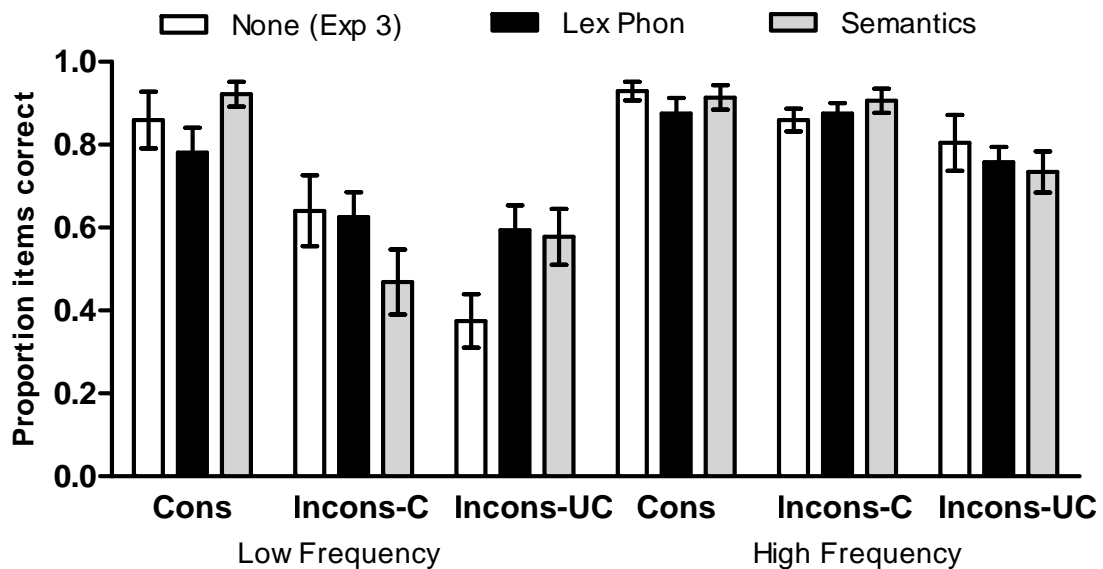


Figure 5.4. Accuracy at the end of training (\pm SE) as a function of pre-exposure, frequency, and consistency.

There was no main effect of pre-exposure, $F_s < 1$. As in Block 1, accuracy was higher for items containing high frequency vowels than for those containing low frequency vowels, $F_p(1, 45) = 45.24, p < .001, \eta_p^2 = .50, F_i(1, 66) = 42.89, p < .001, \eta_p^2 = .39$. Accuracy was also highest for consistent items, followed by inconsistent-conditioned items and was lowest for inconsistent-unconditioned items, $F_p(2, 90) = 31.77, p < .001, \eta_p^2 = .41, F_i(2, 66) = 20.69, p < .001, \eta_p^2 = .39$. The main effects of

frequency and consistency were qualified by a significant interaction, $F_p(2, 90) = 8.60, p < .001, \eta_p^2 = .16, F_i(2, 66) = 6.13, p < .01, \eta_p^2 = .16$. Pairwise comparisons indicated that the effect of frequency was only significant within inconsistent items. In addition, within items containing low frequency vowels, accuracy was higher for consistent items than for both types of inconsistent item which did not differ from each other. Within items containing high frequency vowels, however, accuracy was equally high for consistent and inconsistent-conditioned items and both outperformed inconsistent-unconditioned items.

There was no interaction between pre-exposure and frequency, $F_p < 1, F_i(2, 132) = 1.30, ns$. The interaction between pre-exposure and consistency was not significant by-subjects, $F_p(4, 90) = 1.96, p = .11, \eta_p^2 = .08$, but was by-items, $F_i(4, 132) = 3.74, p < .01, \eta_p^2 = .10$, and the three-way interaction between pre-exposure, frequency, and consistency was significant both by-subjects, $F_p(4, 90) = 3.07, p < .05, \eta_p^2 = .12$, and items, $F_i(4, 132) = 6.41, p < .001, \eta_p^2 = .16$. Pairwise comparisons across both subjects and items demonstrated that within items containing low frequency inconsistent-unconditioned vowels, accuracy was higher in the lexical phonology and semantic conditions than in the no pre-exposure condition. Pairwise comparisons across items also indicated that within items containing low frequency inconsistent-conditioned vowels, accuracy was lower in the semantic condition than in the lexical phonology and no pre-exposure conditions. This effect was not reliable in pairwise comparisons across subjects.

Summary of training performance. At the beginning of training, pre-exposure to either lexical phonology or semantics improved performance across item types relative to the no pre-exposure condition provided by Day 1 of Experiment 3. There was some evidence that semantic pre-exposure was less beneficial than lexical phonology pre-exposure when reading inconsistent-conditioned items. At the end of training, the beneficial effect of lexical phonology and semantic pre-exposure was restricted to items containing low frequency inconsistent-unconditioned vowels. Surprisingly, at the end of training, performance on items containing low frequency inconsistent-conditioned vowels was poorer following semantic pre-exposure than following lexical phonology or no pre-exposure, although this effect was not reliable across subjects.

Old-new decision

Trials with RTs more than two standard deviations away from the mean for that participant were excluded from the analysis (4.6%). Table 5.1 shows accuracy and RTs to trained and untrained items following pre-exposure to lexical phonology and semantics. It also presents results from one sample t-tests which demonstrated that accuracy was above chance for trained and untrained items in both pre-exposure conditions. Accuracy appears to be somewhat higher in the lexical phonology than the semantic condition. However, RTs seem to be slightly faster in the semantic condition.

Table 5.1

Old-new decision accuracy and RTs to trained and untrained items and one-sample t-tests to confirm accuracy above chance.

Item type	Response measure	<i>M</i>	<i>SD</i>	<i>t</i> (15)	<i>p</i>
<u>Lexical Phonology</u>					
Trained	Accuracy	.83	.09	14.12	< .001
	RT (ms)	3436	821		
Untrained	Accuracy	.67	.17	3.91	.001
	RT (ms)	4184	1004		
<u>Semantics</u>					
Trained	Accuracy	.74	.15	6.17	< .001
	RT (ms)	3128	955		
Untrained	Accuracy	.64	.12	4.64	<.01
	RT (ms)	3945	1622		

Masked priming effects in old-new decision. Two analyses of variance were conducted to assess the effects of masked priming on old-new decision performance; one with proportion of items correct as the dependent variable and one using mean RTs to correct items. In the subjects analysis, training status (trained vs. untrained) and prime type (identity vs. control) were treated as within subject variables and pre-exposure (lexical phonology vs. semantics) as a between subjects variable. In the items

analysis, prime type and pre-exposure were within item variables and training status was a between items variable. There were two missing data points in the RT items analysis, one was due to one untrained item receiving no correct responses following identity primes in the lexical phonology condition. The second was due to an untrained item receiving no correct responses following identity primes in the semantic condition. These missing data points were replaced with the mean RT for that item type.

The main effect of prime type was not significant in either the accuracy or RT analysis, $F_s < 1$. There was no interaction between prime type and training status, $F_s < 1$, or prime type and pre-exposure, $F_s < 1$, and no three-way interaction between prime type, training status, and pre-exposure, $F_s < 1$. Thus, masked identity priming had no effect on old-new decision performance.

Trained items were responded to more accurately, $F_p(1, 30) = 15.29, p < .001, \eta_p^2 = .34$, $F_i(1, 46) = 7.52, p < .01, \eta_p^2 = .14$, and faster, $F_p(1, 30) = 45.20, p < .001, \eta_p^2 = .20$, $F_i(1, 46) = 24.10, p < .001, \eta_p^2 = .34$, than untrained items. This effect of training status did not interact with pre-exposure in either the accuracy, $F_p(1, 30) = 1.04, ns, F_i(1, 46) = 1.73, ns$, or RT analysis, $F_s < 1$. Finally, accuracy was higher in the lexical phonology than the semantic condition, although this effect was only significant by-items, $F_i(1, 46) = 3.90, p = .05, \eta_p^2 = .08$, and not by-subjects, $F_p(1, 30) = 2.86, p = .10, \eta_p^2 = .09$. RTs did not differ in the two pre-exposure conditions, $F_s < 1$.

Pre-exposure, frequency, and consistency effects in old-new decision. The effect of pre-exposure to lexical phonology vs. semantics on old-new decision performance,

relative to the no pre-exposure condition provided by Day 1 of Experiment 3, was then assessed with two further analyses of variance, one with proportion of trained items correct as the dependent variable and one using mean RTs to correct trained items. Pre-exposure (none vs. lexical phonology, vs. semantics), frequency and consistency were entered as factors in these analyses. In the semantic and lexical phonology conditions, responses were collapsed across prime-type. This was felt to be justified given the null effect of prime type reported above. There were some missing data points in the RT subjects analysis due to the fact that several participants had no correct responses to a particular item type, for example, one participant responded incorrectly to all items containing low frequency inconsistent-unconditioned vowels. These missing data points (2.4%) were replaced with the mean RT for that item type and pre-exposure condition.

Figures 5.4 and 5.5 summarise the results of these analyses. Examination of Figure 5.5 suggests that accuracy was unaffected by pre-exposure but that RTs were reduced following pre-exposure, particularly for items containing low frequency vowels. Figure 5.6 shows the overall effects of frequency and consistency on old-new decision, collapsed across pre-exposure. Within items containing low frequency vowels, it appears that accuracy was higher, and RTs were faster, for consistent than inconsistent items. In addition, it seems that RTs were faster for items containing high frequency vowels than for items containing low frequency vowels.

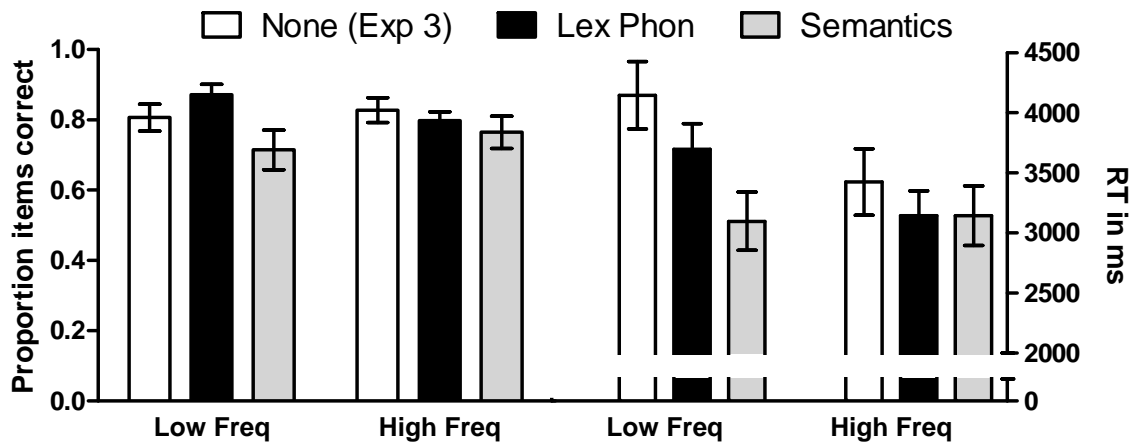


Figure 5.5. Old-new decision accuracy and RTs (\pm SE) as a function of pre-exposure and frequency.

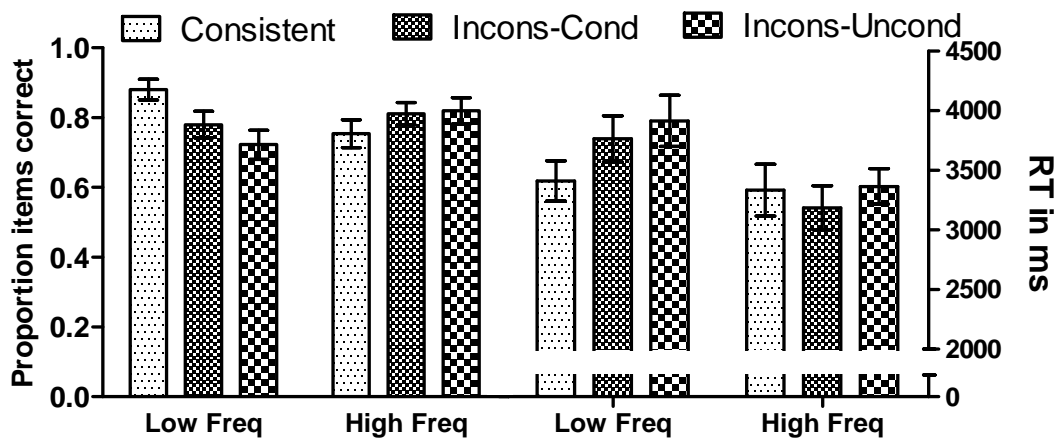


Figure 5.6. Old-new decision accuracy and RTs (\pm SE) as a function of frequency and consistency, collapsed across pre-exposure (none, lexical phonology, semantics).

The analyses revealed that the effect of pre-exposure on accuracy was marginal by-subjects, $F_p(2, 45) = 2.75$, $p = .08$, $\eta_p^2 = .11$, and significant by-items, $F_i(2, 36) = 2.88$, $p < .05$, $\eta_p^2 = .18$. Pairwise comparisons across items indicated that accuracy in the semantic condition was lower than in the lexical phonology condition and marginally lower than in the no pre-exposure condition. Accuracy in the lexical phonology and no

pre-exposure conditions did not differ. The effect of pre-exposure on accuracy did not interact with frequency, $F_p(1, 45) = 1.20$, *ns*, $F_i(2, 36) = 1.63$, *ns*, or consistency, $F_p < 1$, $F_i(2, 36) = 1.77$, $p = .16$, $\eta_p^2 = .17$, and there was no three-way interaction between pre-exposure, frequency, and consistency, $F_p(4, 90) = 1.22$, *ns*, $F_i < 1$.

Turning to the effects of pre-exposure in the latency analysis, RTs were faster in both the lexical phonology and semantic conditions than in the no pre-exposure condition, although this effect was only significant by-items, $F_i(2, 36) = 6.22$, $p < .01$, $\eta_p^2 = .26$, $F_p(2, 45) = 1.58$, *ns*. The effect of pre-exposure interacted with frequency, $F_p(2, 45) = 3.57$, $p < .05$, $\eta_p^2 = .14$, $F_i(2, 36) = 3.19$, $p = .05$, $\eta_p^2 = .15$. Pairwise comparisons indicated that pre-exposure reduced RTs for items containing low but not high frequency vowels, and that RTs were faster to items containing high frequency vowels than to items containing low frequency vowels in the no pre-exposure and lexical phonology conditions, but not in the semantic condition. There was no interaction between pre-exposure and consistency, $F_s < 1$, and no three-way interaction between pre-exposure, frequency, and consistency, $F_p(4, 90) = 1.25$, *ns*, $F_i(4, 36) = 1.27$, *ns*.

Considering the overall effects of frequency and consistency (summarised in Figure 5.6), in the accuracy analysis there was no main effect of frequency, $F_s < 1$, or consistency, $F_s < 1$. However, there was an interaction between frequency and consistency that was significant by-subjects, $F_p(2, 90) = 6.32$, $p < .01$, $\eta_p^2 = .12$, and marginal by-items, $F_i(2, 18) = 2.83$, $p = .09$, $\eta_p^2 = .24$. Pairwise comparisons indicated that there was no effect of consistency within items containing high frequency vowels, but

that within items containing low frequency vowels, accuracy was higher for consistent than inconsistent items. RTs were faster to items containing high frequency vowels than to those containing low frequency vowels, $F_p(1, 45) = 11.40, p < .01, \eta_p^2 = .20, F_l(1, 18) = 6.84, p < .05, \eta_p^2 = .28$, although as stated earlier this was only the case in the no pre-exposure and lexical phonology conditions. In the latency analysis there was no effect of consistency, $F_p(2, 90) = 1.62, ns, F_l < 1$, and no interaction between frequency and consistency, $F_p(2, 90) = 2.05, p = .14, \eta_p^2 = .14$.

Generalization

Figure 5.7 shows generalization accuracy following pre-exposure to lexical phonology vs. semantics, relative to the no pre-exposure condition provided by Day 1 of Experiment 3. Accuracy in both pre-exposure conditions was lower than in the no pre-exposure condition, $F_p(2, 45) = 4.32, p < .05, \eta_p^2 = .16, F_l(2, 36) = 4.87, p = .01, \eta_p^2 = .21$. Pre-exposure did not interact with consistency, $F_p < 1, F_l(4, 36) = 1.19, ns$, or frequency, $F_p(2, 45) = 1.13, ns, F_l(2, 36) = 1.37, ns$, and the three-way interaction between pre-exposure, frequency, and consistency was also non-significant, $F_s < 1$. However, it is clear from Figure 5.7 that the decreased accuracy in the pre-exposure conditions was more pronounced for items containing inconsistent vowels.

Overall, performance was better for items containing high frequency vowels, $F_p(1, 45) = 34.96, p < .001, \eta_p^2 = .44, F_l(1, 18) = 82.64, p < .001, \eta_p^2 = .82$, and for items containing consistent vowels than for those containing inconsistent vowels, $F_p(2, 90) =$

15.44, $p < .001$, $\eta_p^2 = .26$, $F_i(2, 18) = 42.25$, $p < .001$, $\eta_p^2 = .82$. The consistency \times frequency interaction was also significant, $F_p(2, 90) = 3.44$, $p < .05$, $\eta_p^2 = .07$, $F_i(2, 18) = 7.10$, $p < .01$, $\eta_p^2 = .44$. Pairwise comparisons and examination of effects sizes indicated that the frequency effect was significant across consistency levels, but was strongest in items containing inconsistent-unconditioned vowels, followed by items containing inconsistent-conditioned vowels, and weakest in items containing consistent vowels. The consistency effect was also significant and took the same form across frequency levels, but was stronger in items containing low frequency vowels than in items containing high frequency vowels.

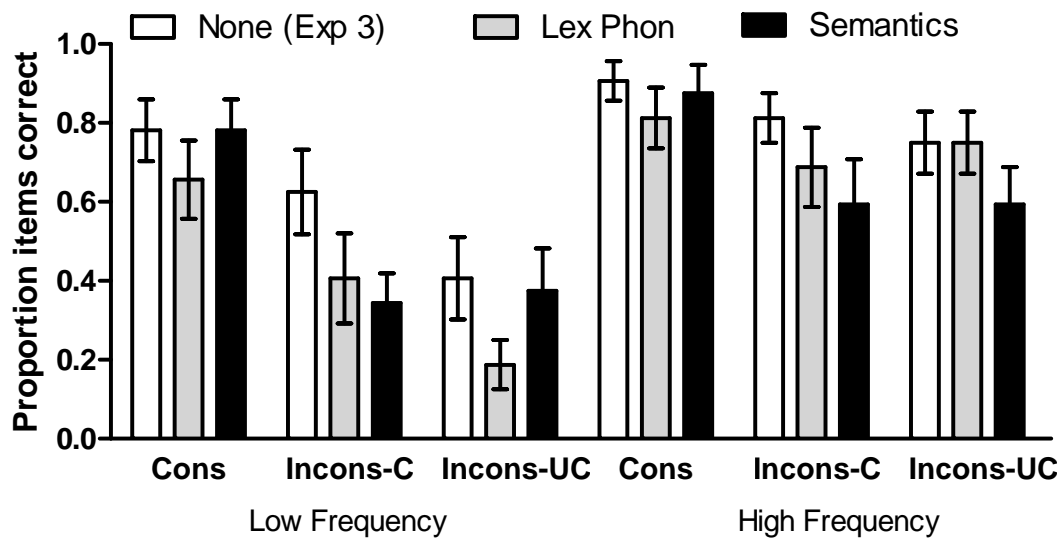


Figure 5.7. Generalization accuracy (\pm SE) as a function of pre-exposure, frequency, and consistency.

Discussion

Rationale

This experiment investigated the hypothesis that semantics supports learning to read, as embodied in the triangle model of reading aloud (Harm & Seidenberg, 2004; Plaut et al., 1996). Previous research on typical adults (Balota et al., 2004; Cortese & Khanna, 2007; Frost et al., 2005; Pexman et al., 2008; Strain et al., 1995) and children (Ricketts et al., 2007), and neuropsychological patients (McKay et al., 2007; Woollams et al., 2007), supports the view that semantics plays a role in reading aloud and visual word recognition. However, such research has suffered from several methodological problems; including the inadequacy of lexical database estimates of variables such as frequency, consistency, and imageability, the use of inappropriate experimental designs in which graded variables are artificially categorised, the difficulty of operationalizing semantics, and the failure to control for potentially important and correlated variables. Furthermore, previous research leaves open the possibility that the purported influence of semantics on orthographic processing may in fact be driven by familiarity with word sounds rather than familiarity with word meanings.

McKay et al. (2008) attempted to overcome some of these problems and address the issue of phonological vs. semantic familiarity using a novel word learning methodology. However, this study again suffered from several methodological problems. First, it is difficult to ascertain whether phonological familiarity provided any benefit for subsequent learning because semantic vs. phonological familiarity was not compared to a baseline no familiarity condition. Second, because pre-exposure was

manipulated within subject, learning definitions during semantic pre-exposure had the potential to interfere with attending to item sounds during lexical phonology pre-exposure. Third, novel words were written in English orthography meaning that reading accuracy was extremely high from the beginning of training. Thus, McKay et al.'s experiment did not in fact address the question of whether semantics plays a role in *learning to read*.

The current experiment sought to address these issues utilising the control provided by the artificial orthography paradigm introduced in Experiments 2 and 3. Thirty-two adults learned to read the artificial orthography to 70% accuracy. Prior to this, half the participants were pre-exposed to the sounds of the items (lexical phonology) and half to their sounds plus a novel object referent (semantics). This between subjects pre-exposure manipulation meant that learning semantic associations did not have the potential to interfere with attending to lexical phonology item sounds. Attention during pre-exposure was also ensured by including a monitoring task in which participants were intermittently asked whether a particular item had been repeated in the previous five trials. Performance following pre-exposure was compared to that on Day 1 of Experiment 3 in which participants received no pre-exposure prior to orthography training. This enabled direct examination of any benefits obtained from phonological familiarity alone. A masked priming manipulation was included in old-new decision to assess the extent to which trained items were treated in a similar way to words in natural languages.

The fact that novel words were written in novel orthography means that the current experiment addressed the role of semantics in *learning* to read. The method also avoids problems associated with extracting variables from lexical databases, categorizing graded variables, and the influence of non-controlled factors such as AoA. Furthermore, Experiments 2 and 3 established that strong, reliable, and predictable effects of spelling-sound frequency and consistency emerge when learning to read the novel orthography. This makes it an ideal paradigm for investigating whether semantic familiarity has a greater impact when reading aloud low frequency inconsistent words, as suggested by the triangle model.

Predictions

It was predicted that pre-exposure to semantics would boost subsequent reading accuracy for trained items containing low frequency inconsistent vowels. Semantic pre-exposure was also expected to decrease error rates and RTs in old-new decision. Previous research suggested that lexical phonology pre-exposure might also improve orthographic learning. However, it seemed likely that lexical phonology pre-exposure would provide less benefit than semantic pre-exposure. In old-new decision, RTs to trained (but not untrained) items should have been decreased by prior presentation of an identity prime relative to a control prime. Masked priming effects were not expected to differ in the two pre-exposure conditions. No explicit predictions were made about the influence of pre-exposure on generalization.

Summary of results

At the beginning of orthography training, pre-exposure to semantics increased reading accuracy across item types, relative to the no pre-exposure condition provided by Day 1 of Experiment 3. By the end of training, this facilitation was specific to items containing low frequency inconsistent-unconditioned vowels. Relative to the no pre-exposure baseline, semantic pre-exposure also decreased old-new decision latencies to trained items, particularly when they contained low frequency vowels. Generalization performance was in fact poorer following pre-exposure to semantics than in the no pre-exposure condition provided by Day 1 of Experiment 3. Importantly, none of these effects were specific to the semantic condition: reading accuracy and old-new decision RTs were similarly boosted (and generalization similarly depressed) by pre-exposure to lexical phonology alone. Pre-exposure did not improve old-new decision accuracy, and masked identity priming did not influence old-new decision.

The following paragraphs first consider the effects of phonological vs. semantic familiarity on performance at the end of orthography training. The difference between the effects of pre-exposure early and late in training are then explored. Next, discrimination and generalization results are interpreted, followed by a discussion of the extent to which trained items were treated similarly to real words. Finally, the validity of representing semantics with novel objects is considered in some detail.

The effect of semantic vs. phonological familiarity on end of training performance

At the end of training, pre-exposure to semantics increased reading accuracy for trained items containing low frequency inconsistent-unconditioned vowels, relative to

the no pre-exposure baseline provided by Day 1 of Experiment 3. This might be taken as support for the idea that semantics plays a role in reading aloud. However, pre-exposure to lexical phonology showed exactly the same effect, suggesting that supposed semantic effects on reading aloud may be driven by familiarity with item sounds rather than familiarity with item meanings. It is important to note that the current experiment did not obtain a null effect of pre-exposure: pre-exposure to *both* lexical phonology *and* semantics improved subsequent reading accuracy for items with low frequency inconsistent spelling sound mappings, relative to no pre-exposure.

This finding contrasts with those reported by McKay et al. (2008) in which pre-exposure to semantics (via learning definitions) enhanced reading of inconsistent novel words relative to pre-exposure to lexical phonology (repeating the phonological forms of items). In their paradigm, the two types of pre-exposure were manipulated within subject. Potentially therefore, the onerous task of learning definitions in the semantic condition interfered with attending to the sounds of the items in the lexical phonology condition. As pre-exposure was manipulated between subjects in the current experiment, there was no risk of interference between conditions. The present findings therefore suggest that McKay et al. may have underestimated the benefit of phonological familiarity, and thus over-estimated the benefit of semantic familiarity, for subsequent orthographic processing. This leaves open the possibility that what supports inconsistent word reading is familiarity with the overall sound of a word, rather than familiarity with its meaning.

A further finding deserves some consideration. At the end of orthography training, reading accuracy for items containing low frequency inconsistent-conditioned vowels was in fact lower in the semantic condition than in the lexical phonology and no pre-exposure conditions. Inconsistent-conditioned items were characterised by a repeated onset-vowel combination. Experiments 2 and 3 indicated that this predictability was helpful for learning to read items aloud: conditioned inconsistent items reliably outperformed unconditioned inconsistent items. One possibility is that participants in the semantic condition concentrated their attention on learning the object-label associations and that this reduced their sensitivity to the predictable onset-vowels in inconsistent-conditioned items.

This explanation is entirely speculative and it is perhaps unwise to place too much emphasis on this unexpected finding. Although predictable consonant-vowel relationships do characterise English orthography, they are less salient than those in the artificial orthography. Thus, this effect is potentially an artefact of the restricted learning environment of the current experiment. However, it should be re-emphasised that although consonant-vowel pronunciation rules in the artificial orthography were entirely deterministic and some were fairly salient, the majority of participants were not aware of these rules. In the current experiment, post-experiment debriefing indicated that one participant had noticed the onset-vowel relationship that characterised items containing high frequency inconsistent-conditioned vowels. However, no other participants had explicit knowledge of conditional vowel pronunciation rules.

The effect of pre-exposure earlier in training

At the beginning of orthography training, pre-exposure to semantics or lexical phonology increased reading accuracy, relative to no pre-exposure, across item types. This suggests that facilitative effects of phonological/semantic familiarity may change over time, with the specificity of effects to low frequency inconsistent items emerging with increasing proficiency. Previous research has not addressed this possibility because proficiency has been high enough for consistent items to be at, or near, ceiling. For example, in Ricketts et al.'s (2007) study, 9-year-old children read 90% of a list of 30 regular words correctly, and in McKay et al.'s (2008) experiment, consistent novel word reading accuracy was at ceiling from the beginning of training. Even when latencies as opposed to errors are examined, as in Strain et al. (1995), RTs for inconsistent words are typically more variable and therefore more likely to show semantic effects. The results of the current study remind us that the effect of semantics, and indeed the distinction between consistent and inconsistent words, should not be thought of as all-or-nothing and suggest that the effect of semantic/phonological familiarity on reading aloud may become restricted to particular item types over time.

Unfortunately, the developmental relationship between semantics and reading aloud has been neglected in simulations of the triangle model. Plaut et al. (1996) made the assumption that semantics would only make a substantial contribution to reading aloud once orthography-phonology mappings had developed to some degree. They suggested that this should be the case "... in part because of the phonological nature of typical reading instruction and in part because in English, the orthography-to-phonology

mapping is far more structured than the orthography-to-semantics mapping...” (p. 95). On this basis, Plaut et al. chose to increase the input from semantics over the course of training, meaning that there was no opportunity for semantics to influence the earliest stages of learning to read. Whilst the above points are valid, Plaut et al. also acknowledge that “... the mapping between semantics and phonology develops, in large part, prior to reading acquisition...” (p. 95). It therefore seems equally likely that semantics might play a role in reading aloud from the outset and it is disappointing that this was not explored in their simulations.

Some support for this suggestion comes from an experiment by Laing and Hulme (1999) in which they taught 4- to 6-year-old children to associate three or four letter printed cues/abbreviations with spoken word targets, e.g., *bfr* represented “*beaver*”. Children’s ability to learn the associations was influenced by the phonological similarity of the cue to the target, as well as the imageability of the target. Importantly, however, no interaction between these two variables was observed. The results of this experiment suggest that, in the early stages of reading development, semantic variables impact on the mapping of orthography to phonology irrespective of the ease of this mapping.

More direct evidence is provided by two recent developmental studies. Nilsen and Bourassa (2008) taught 6-year-old children to read new words which had either regular or irregular spelling-sound patterns, e.g., SNAKE vs. SWORD. The novel words had either concrete or abstract meanings, e.g., ELBOW vs. TEMPER. Children learned to read regular words more easily than irregular words, and concrete words more easily

than abstract words. Again, however, no regularity by meaning interaction emerged; concrete words were learned more easily irrespective of the regularity of their spelling-sound mappings. This suggests that in the earliest stages of learning to read, semantics may support the learning of orthography-phonology mappings for all words.

Ouelette and Fraser (in press) also found that familiarity with word meanings was beneficial when learning to read words with regular and consistent spelling-sound mappings. Children aged 9.5 years learned to read novel words, such as YAIT or ROOP. Five were learned in isolation (non-semantic), and five were learned in combination with a colour picture of the word and a short orally presented passage describing its meaning (semantic). To control for phonological exposure in the semantic and non-semantic conditions, the novel word itself did not occur in the passage. Words learned in conjunction with meanings were more likely to be identified correctly in a four alternative orthographic choice task, both one day and four days after training. Although this experiment did not train any of the novel words with irregular/inconsistent spelling-sound mappings, the results do suggest that knowledge of word meanings aids orthographic learning for regular/consistent words, even in mid-childhood.

It should be emphasised that none of the above studies ensured that attention was equivalent across conditions or explicitly examined the contributions of knowledge of word meanings vs. knowledge of word sounds. For example, the effect of semantics seen in Ouelette and Fraser's (in press) experiment may have resulted from children paying more attention to words that were given definitions, rather than from the semantic content of the definitions. Thus, these studies support the idea that, early in

learning, semantic/phonological familiarity influences reading aloud irrespective of spelling-sound regularity, but do not provide strong evidence for an influence of semantics over and above effects of phonological familiarity.

Discrimination

In Experiment 3 it was suggested that the lack of meaning in the artificial orthography might have contributed to some of the differences between old-new decision in the artificial orthography and lexical decision in natural languages. This is supported by natural language experiments which show that semantic variables influence lexical decision (Balota et al., 2004; Cortese & Khanna, 2007; Pexman et al., 2008), and the novel word learning study of McKay et al. (2008), which found that discrimination of trained from untrained items was improved by semantic pre-exposure. In the current experiment, old-new decisions were faster in the semantic condition than in the no pre-exposure condition, within items containing low but not high frequency vowels. However, exactly the same facilitation was seen following pre-exposure to lexical phonology. As with the reading aloud data, it seems likely that the within subject pre-exposure manipulation used by McKay et al. underestimated the benefit of phonological familiarity for visual word recognition.

Unexpectedly, old-new decision accuracy was lower in the semantic condition than the lexical phonology condition and in fact, neither condition differed significantly from the no pre-exposure baseline. The fact that the accuracy and RT data do not pattern together suggests that we should not place too much weight on the discrimination data. Nevertheless, there is no evidence that semantic pre-exposure

improved discrimination over and above lexical phonology pre-exposure; with the benefits of pre-exposure not being clear in either condition.

As in Experiment 3, discrimination was influenced in predictable ways by the frequency and consistency characteristics of trained items. Within items containing low frequency vowels, accuracy was higher when they were consistent than when they were inconsistent. There was no overall effect of frequency on accuracy, but RTs were faster to trained items containing high frequency vowels than to those containing low frequency vowels. Consistency did not affect RTs. Thus, it has now been shown across three experiments that the effects of frequency and consistency on both reading aloud and discrimination in the artificial orthography are similar to those observed in English orthography.

Generalization

Faster discrimination following pre-exposure brought with it a cost. Although generalization showed the same effects of frequency and consistency as on Day 1 of Experiment 3, overall accuracy was lower following pre-exposure to lexical phonology or semantics. Why might pre-exposure reduce generalization? One possibility is that because pre-exposure resulted in stronger representations of the sounds of trained items, this enabled participants to read training items by recalling that they were a valid word in the language. In turn this may have reduced the necessity to extract information about the influence of consonants on vowel pronunciations, resulting in reduced generalization. In support of this, the reduction in generalization performance was particularly clear for items containing context-sensitive inconsistent vowels.

Previous research supports the view that any manipulation that reduces attention to character-sound mappings might be detrimental for generalization. For example, Harm, McCandliss, and Seidenberg (2003) found that, in a connectionist model, training that emphasised sub-word regularities between orthography and phonology produced greater benefit for nonword reading (generalization) than training which focused only on phonology. This demonstrates the importance of focusing on spelling-sound relationships for the development of generalization. More direct evidence is provided by Landi, Perfetti, Bolger, Dunlap, and Foorman (2006). They found that children were better at learning to read new words presented in connected text than new words presented in isolation. However, when re-tested one week later, reading accuracy was higher for those words learned in isolation. This suggests that although context increased the chance of reading words correctly during training, it served to reduce attention to orthography-phonology mappings, which in turn compromised longer term retention. Relating these findings to the current experiment, pre-exposure may have provided an additional source of information to support reading acquisition of trained items, while simultaneously reducing the extraction of conditional character-sound mappings. Landi et al. also noted that learning in context had a less negative effect on retention for those children who were more advanced readers. This suggests that pre-exposure would be less detrimental for generalization if orthography training was more extensive.

Were trained items processed similarly to real words?

As described earlier, although pre-exposure decreased the time participants took to discriminate trained from untrained items, it did not improve their accuracy. In fact, accuracy in the semantic condition was lower than in the lexical phonology condition and slightly lower than in the no pre-exposure condition, although this effect was only marginal across subjects. The fact that different effects were seen in errors and RTs alludes to the possibility that old-new decision in the artificial orthography is not directly comparable to McKay et al.'s (2008) old-new decision task, or lexical decision in standard English. In both these cases, the effect of semantic variables on error rates mirrors the facilitation seen in latencies.

This position is bolstered by the fact that discrimination of trained from untrained items was unaffected by a masked priming manipulation. In contrast, in lexical decision in natural languages, words (but not nonwords) are responded to faster when preceded by an identity prime than when preceded by a control prime (Bowers, 2003; Davis et al., 2008; Forster et al., 2003; Forster, 1998; Forster & Davis, 1984). It is possible that the discrimination task lacked power to detect priming effects. In addition, the laborious nature of old-new decision in the artificial orthography may have reduced the task's sensitivity to the priming manipulation. Both these suggestions could be tested by considerably extending the length of orthography training in order to increase the automaticity of recognition, and by including more items in old-new decision. However, at present, results suggest that trained items were not processed in entirely the same way as words in natural languages.

The long old-new decision latencies suggest that participants were decoding items and using a combination of orthographic and phonological familiarity to consciously evaluate whether they recognised them. Thus, unlike in typical lexical decision, recognition of trained items was not automatic and discrimination was rather strategic. If trained items are not processed in the same way as words in natural languages, perhaps effects seen in the artificial orthography do not inform us about lexical processing in natural languages? However, in both the current and preceding experiments, lexical-like effects of frequency and consistency were present during orthographic learning, discrimination, and generalization. Furthermore, pre-exposure boosted reading accuracy for exactly those items that the triangle model and existing research predicts should be sensitive to semantic variables. Thus, although there are differences between trained and real word item processing, the artificial orthography paradigm provides a level of experimental control that is difficult for natural language experiments to obtain and generates data that mirror typical reading processes. There does not therefore seem good reason reject it as a method for investigating the factors influencing orthographic learning.

Are object-label associations really semantic?

An important qualification to the argument that semantic familiarity does not support orthographic learning beyond the contribution of phonological familiarity, is that semantics were instantiated using novel object referents. As argued in the introduction to this experiment, learning object-label associations is a crucial component of word learning (Estes et al., 2007; Werker et al., 1998) and participants in

the current experiment learned the object-label associations well. However, it seems likely that the semantic representations that were created would not have been as strong or rich as those that characterise words in natural languages, or indeed those created by the novel definitions used by McKay et al. (2008). Potentially, therefore, they may have been too weak to drive semantic influences on reading aloud and discrimination, over and above the effects induced by phonological familiarity. On this view, had semantics been instantiated with definitions, the facilitation seen in the semantic condition might have been greater. This possibility is explored in Experiment 5.

Summary and conclusions

In this experiment, phonological familiarity improved reading accuracy during training. By the end of orthography training, this facilitation was specific to items containing low frequency inconsistent-unconditioned vowels. Phonological familiarity also increased discrimination speed, but not accuracy. Surprisingly, facilitation was no greater in a semantic condition in which participants were familiarised with associations between the phonological form of items and a novel object referent. These findings suggest that the role of semantics in learning to read, embodied in the triangle model, might be better recast as an effect of possessing item-specific phonological representations.

Old-new decision was not influenced by a masked priming manipulation and latencies were still extremely long. This suggests that trained items are not represented in the same way as words in natural languages. However, multiple experiments have now demonstrated that learning, discrimination, and generalization in the artificial

orthography are similarly influenced by lexical variables as words in natural languages. Together with the high degree of experimental control provided by the paradigm, it still seems a valid and useful tool for investigating the factors influencing orthographic learning.

Generalization performance was depressed following pre-exposure, relative to the no pre-exposure condition. Plausibly, pre-exposure allowed participants to read training items by recalling that they were a valid word in the language and this may have reduced the necessity of extracting information about the influence of consonants on vowel pronunciation, essential for successful generalization. A study by Landi et al. (2006) described earlier suggests that increased orthography training would reduce the detrimental effects of pre-exposure.

An important caveat to the argument that phonological rather than semantic familiarity supports orthographic learning is that novel object referents may have induced only weak semantic representations. These may not have been strong enough to support semantic effects on reading aloud and visual word recognition, over and above effects of phonological familiarity. This possibility is the main focus of Experiment 5, in which semantic associations are instantiated using novel definitions as opposed to novel objects.

CHAPTER 6: IS PHONOLOGICAL OR SEMANTIC FAMILIARITY IMPORTANT IN LEARNING TO READ?

This experiment explored whether learning to read new words in an artificial orthography benefits from familiarity with word meanings, instantiated using definitions. In Experiment 4, phonological familiarity provided equivalent benefit to object-label associations, but these may have induced only weak semantic representations. In the current experiment, 32 participants received definitions pre-exposure for half the items and lexical phonology pre-exposure for the remaining items. This within subject manipulation meant that learning semantic item definitions might reduce attention to lexical phonology item sounds, therefore performance was also compared to the (between subjects) lexical phonology condition from Experiment 4. Cross-experiment analyses showed that semantic pre-exposure benefited orthographic learning more than lexical phonology pre-exposure. However, lexical phonology and semantic pre-exposure were equally beneficial by the end of training. Extended orthography training and richer semantic representations might induce an end of training semantic advantage.

Previous natural language research suggests that semantics influences reading aloud, particularly when words are low in frequency and have inconsistent spelling-sound mappings (Balota et al., 2004; Cortese & Khanna, 2007; Frost et al., 2005; McKay

et al., 2007; Patterson et al., 2006; Strain et al., 1995; Woollams et al., 2007). Semantic variables also affect visual word recognition, as indexed by lexical decision (Balota et al., 2004; Cortese & Khanna, 2007; Pexman et al., 2008). This role for semantics in orthographic processing is embodied in the triangle model (Harm & Seidenberg, 2004; Plaut et al., 1996). However, as discussed in preceding Chapters, there are several methodological problems that limit the conclusions that can be drawn from the existing literature.

Particularly problematic is the natural correlation between familiarity with word meanings and familiarity with word sounds. To re-iterate some examples provided in Chapter 5, age-of-acquisition (AoA) has been shown to account for imageability effects in reading aloud (Cortese & Khanna, 2007; Ellis & Monaghan, 2002; Monaghan & Ellis, 2002), and early acquired words have stronger phonological as well as semantic representations. Thus, it follows that imageability effects, traditionally thought to index semantic factors, may be driven by phonological rather than semantic familiarity. As a second example, oral vocabulary knowledge is related to exception word reading in children (Ricketts et al., 2007), supporting a link between semantics and reading aloud that is strongest for words with atypical spelling-sound mappings. But, the better a child understands a word's meaning, the more familiar they are likely to be with how it sounds. Thus, the relationship between oral vocabulary and exception word reading could again be driven by phonological not semantic familiarity. Natural language studies do not, therefore, provide unequivocal evidence that semantics plays a direct role in

orthographic processing and leave open the possibility that phonological familiarity is the crucial variable.

A recent novel word learning study by McKay et al. (2008) overcame some of these methodological problems and investigated the role of semantic vs. phonological familiarity in learning to read novel words written in English orthography. Pre-exposure to definitions of novel words (semantics) provided greater benefit for subsequent orthographic processing than pre-exposure to novel word sounds (lexical phonology). In reading aloud, this benefit was specific to novel words that had been assigned inconsistent spelling-sound mappings; in old-new decision, the facilitation was evident across items.

As discussed in Chapter 5, however, there are several issues with McKay et al.'s (2008) experiment which weaken the strength of their findings. First, semantic vs. lexical phonology pre-exposure was not compared to a no pre-exposure baseline. It is therefore possible that phonological familiarity did benefit subsequent orthographic processing, but to a lesser extent than semantic familiarity. Second, the use of a within subject semantic manipulation may have reduced the benefit of lexical phonology pre-exposure. To elaborate, during pre-exposure, participants learned the definitions of semantic items and this task was interspersed with blocks in which they listened to and repeated the sounds of lexical phonology items. It is therefore possible that participants concentrated on learning the definitions and paid minimal attention during listen-repeat trials. Thus, semantic and phonological familiarity may provide equal benefit in a situation which promotes full attention in both conditions.

A further issue, not addressed by McKay et al.'s (2008) experiment or the natural language studies cited earlier, is the role of semantics in the early stages of orthographic learning. Natural language studies have either focused on skilled adult readers or relatively competent child readers and although McKay et al. used novel words, they were written in English orthography and reading accuracy was high from the beginning of training. Thus, existing research does not inform as to whether the role of semantics in orthographic processing changes over time.

Experiment 4 investigated the impact of semantic vs. phonological familiarity on learning to read using the artificial orthography paradigm introduced in Experiment 2. Prior to orthography training, 16 participants were familiarised with the phonological forms of novel items (lexical phonology pre-exposure), and 16 participants were familiarised with the phonological forms of items plus a novel object referent (semantic pre-exposure). This between subjects pre-exposure manipulation ensured that learning in the semantic condition could not interfere with learning in the lexical phonology condition. Furthermore, results were compared to a no familiarity condition (Day 1, Experiment 3), providing a baseline against which to evaluate any effects of semantic vs. lexical phonology pre-exposure. Because the orthography was novel, the effect of pre-exposure could be assessed in early and later stages of learning. The artificial orthography paradigm also gave complete control over exposure to the frequency and consistency characteristics of the language, something that is difficult to achieve when using natural alphabetic scripts.

Experiment 4 failed to replicate McKay et al.'s (2008) findings. Relative to the no pre-exposure baseline, lexical phonology pre-exposure increased reading accuracy during training to the same extent as semantic pre-exposure. By the end of training this benefit was specific to items containing low frequency inconsistent-unconditioned vowels. Semantic and lexical phonology pre-exposure also increased old-new decision speed to the same extent. Note though that this facilitation was not reflected in the accuracy data, suggesting that some caution should be taken when interpreting old-new decision data. Overall, these findings support the argument that knowledge of word meanings does not support orthographic learning over and above item-specific knowledge of word sounds.

Three further important results emerged from Experiment 4. First, early in training, pre-exposure benefited performance across items, whereas by the end of training, facilitation was restricted to items containing low frequency inconsistent vowels. Thus, effects of semantic/phonological familiarity on reading aloud may differ in early vs. later stages of learning. Second, frequency and consistency had predictable effects on old-new decision performance. Within items containing low frequency vowels, accuracy was higher for consistent than inconsistent items, and, across items, RTs were faster when they contained high frequency vowels. These effects are in the same direction as those reported in typical lexical decision (e.g., Lacruz & Folk, 2004). Note, however, that effects were observed in either accuracy or latencies, whereas in typical lexical decision such effects tend to be obtained across these outcome measures. Third, pre-exposure reduced generalization performance. It was suggested that because

pre-exposure increased knowledge of which items constituted words in the language, this might reduce the need to extract knowledge of the rules that governed inconsistent vowel pronunciations. This would be detrimental for generalization.

Experiment 5

The current experiment sought to explain why the effects of lexical phonology and semantic pre-exposure observed in Experiment 4 differed from those reported by McKay et al. (2008). Also of interest was whether the results of the current experiment would replicate several effects observed in Experiment 4: a) the differential effects of pre-exposure early and late in training, b) frequency and consistency effects in old-new decision, and c) reduced generalization following pre-exposure.

Outline of method

Thirty-two participants learned to read the artificial orthography. Prior to this, they were familiarised with the sounds of all items (*lexical phonology pre-exposure*) and learned novel definitions for half the items (*semantic pre-exposure*). The impact of this within subject pre-exposure manipulation on reading accuracy during training and old-new decision was assessed within the current experiment, but also with reference (between subjects) to the lexical phonology condition from Experiment 4 and the no pre-exposure baseline condition provided by Day 1 of Experiment 3. More items were included in old-new decision. In part, this was to ensure that power was sufficient to examine the impact of the within subject pre-exposure manipulation. However, it was also hoped that this increased power would strengthen the frequency and consistency

effects observed in Experiment 4. The overall impact of pre-exposure on generalization was assessed with reference to the no pre-exposure condition.

Hypotheses and predictions

With regards to the effects of pre-exposure on reading accuracy during training and discrimination, two opposing hypotheses were proposed. Hypothesis 1 was that Experiment 4 failed to find a semantic advantage, over and above the benefit observed in the lexical phonology condition, because novel objects induced weaker semantic representations than the definitions used by McKay et al. (2008). Hypothesis 2 was that the advantage of semantic over lexical phonology pre-exposure observed in McKay et al.'s experiment resulted not from the semantic nature of pre-exposure, but due to reduced benefit of lexical phonology pre-exposure, caused by the within subject manipulation. On this view, participants concentrated on learning the semantic definitions and paid minimal attention during lexical phonology listen-repeat trials.

Both of the above hypotheses predict that in the current experiment, in which pre-exposure was manipulated within subject, semantic pre-exposure should provide greater benefit than lexical phonology pre-exposure, as observed by McKay et al. (2008). However, they make different predictions for a comparison between the current experiment and Experiment 4. If Experiment 4 failed to demonstrate an effect of semantics because the object referent task produced weak semantic representations, the benefit of semantic pre-exposure in the current experiment should also be greater than the benefit of lexical phonology pre-exposure seen in Experiment 4. Alternatively, if a within subject design (used in the current experiment and by McKay et al.) induces an

apparent semantic effect by virtue of dampening performance in the lexical phonology condition, the benefit of semantic pre-exposure in the current experiment should be equivalent to the benefit of lexical phonology pre-exposure observed in Experiment 4. Furthermore, this second hypothesis predicts that lexical phonology pre-exposure should provide relatively less benefit in the current experiment than in Experiment 4.

Three further predictions were made. First, pre-exposure was expected to confer benefit to all items at the beginning of orthography training but only to items containing low frequency inconsistent-unconditioned vowels by the end of training, as in Experiment 4. Second, with more items included in old-new decision, it was hoped that the frequency and consistency effects observed in Experiment 4 would replicate and possibly be stronger. Finally, the current experiment also provided an opportunity to examine whether, as in Experiment 4, pre-exposure would lead to worse generalization, relative to the no pre-exposure condition provided by Day 1 of Experiment 3.

Method

Participants

Thirty-two adults (9 males, 23 females) took part in this experiment. Their mean age was 22.06 years, $SD = 3.24$. All participants had English as a first language and had normal hearing and vision and none had participated in Experiments 2 to 4.

Materials and Procedure

Pre-exposure phase. Participants first completed a pre-exposure phase that involved learning the sounds for all the items (*lexical phonology*) and a definition for half

of the items (*semantics*). The procedure for this pre-exposure phase is summarised in Figure 6.1. Definitions were adapted from Oxford English Dictionary entries of extremely low-frequency concrete English words, see Appendix 6.A. Seventeen of these were taken from McKay et al. (2008), and one, “an assistant to a magician or scholar”, was added. For each participant, 18 items were assigned to the semantic condition and 18 to the lexical phonology condition. Within each condition, 12 items contained high frequency vowels and 6 contained low frequency vowels. Within the sets of high and low frequency items, consistent, inconsistent-conditioned, and inconsistent-unconditioned items were equally represented. Sixteen participants received one half of the items in the lexical phonology condition and the other half in the semantic condition, and the remaining 16 participants received the reverse assignment.

Participants were told that they would be learning how to say some new words and that for half the words they would also be learning their meanings. They were asked to try to remember the meanings of the words and were told that they would be tested on them. On each trial, a fixation cross appeared on screen and participants listened to and repeated one of the training items. In the *lexical phonology* condition this constituted one trial and the next item was then presented. In the *semantic* condition, a written definition then appeared on screen and participants read the definition either silently or out loud, pressing spacebar to move onto the next trial. Participants experienced six semantic trials, followed by six lexical phonology trials; this process was then repeated twice more so that each of the 36 training items had been presented.

This constituted one pre-exposure block. The order in which the 18 lexical phonology and 18 semantic items were presented was randomised within this block.

Three of the above pre-exposure blocks were completed, followed by a definition recall test. Participants heard one of the semantic items and were asked to say its definition out loud. The experimenter recorded their responses. The correct definition then appeared on screen and participants pressed spacebar to move onto the next recall trial. This process was repeated until each of the 18 semantic items had been presented. Item presentation was randomised. To ensure that the 18 lexical phonology items were heard as many times as the semantic items during the pre-exposure phase, participants then completed a series of six serial recall tasks. These involved listening to three lexical phonology items and repeating them back in the correct order. This relatively demanding task ensured attention to the lexical phonology items. Each of the 18 lexical phonology items was presented once and item presentation was randomised.

Following the definition and lexical phonology recall tasks, three more pre-exposure blocks were completed. The definition and lexical phonology recall tasks were then administered for a second time. This completed the pre-exposure phase.

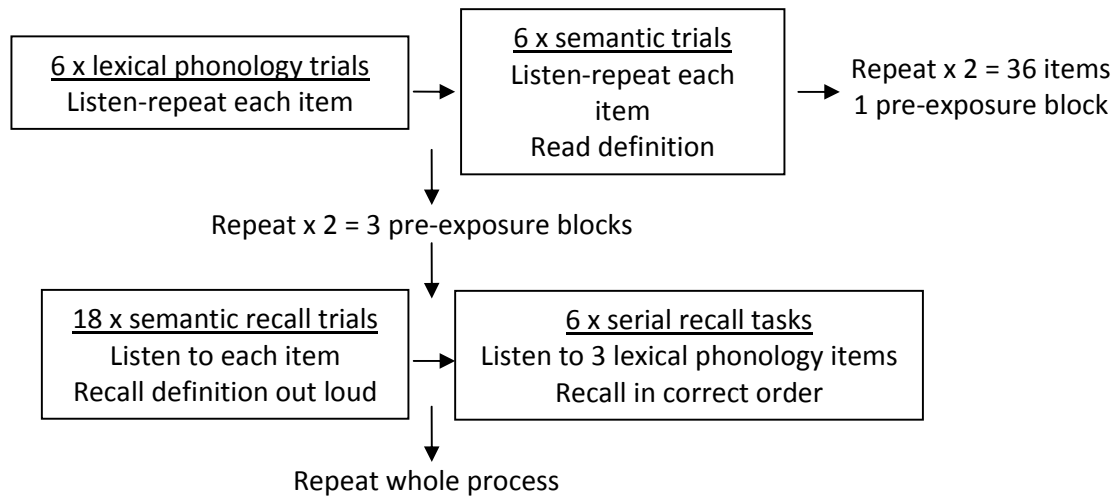


Figure 6.1. Procedure for Experiment 5 pre-exposure phase.

Orthography training and testing. Following pre-exposure, participants learned to read the novel orthography to at least 70% accuracy, according to the procedure described in Experiment 3 - Day 1. Old-new decision was then administered, again as described in Experiment 3 - Day 1. However, all 36 training items were included in old-new decision, along with 24 untrained distractors. These untrained distractors were those used in old-new decision on Day 1 ($n = 12$) and Day 2 ($n = 12$) of Experiment 3. No masked priming manipulation was included, given the null effects obtained in Experiment 4.

After old-new decision, retention of semantic knowledge was checked using a definition recall task. Participants heard one of the semantic items and were asked to recall its definition out loud. Each of the 18 semantic items was presented in a random order. Unlike during pre-exposure, no feedback was given. Generalization was then completed, as described in Experiment 3 - Day 1. The whole procedure took place in a single session lasting around 60 minutes.

Results

Semantic learning

It was first important to ensure that participants had learned and remembered the definitions of the semantic items. A score of 1 was given when the key features of a definition were recalled (e.g., “a wooden case used for storing cannonballs” recalled as “a box for cannonballs”). A score of 0 was given when the wrong or no definition was recalled, or when only minimal features were remembered (e.g., “the fold of skin hanging from the neck of cattle” recalled as “something to do with a cow”). The mean proportion of definitions correctly recalled was the same at the end of pre-exposure, $M = .93$ ($SD = .25$) and following the old-new decision task, $M = .93$ ($SD = .25$), demonstrating that the meanings of the semantic items had been learned and remembered.

Training

Overall performance during training. In the current experiment, the maximum number of blocks required to achieve the criterion of 70% accuracy was 5. The mean number of blocks to achieve criterion was 1.97 ($SD = 1.06$). Two of the analyses reported below compare performance in the current experiment with that in the no pre-exposure condition provided by Day 1 of Experiment 3, in which some participants required 6 blocks to achieve criterion. In order to make these comparisons, when a participant required less than 6 blocks to achieve criterion, accuracy in remaining blocks was assumed to be identical to that in the final block of training completed.

Figure 6.2 shows performance during training following pre-exposure to semantics vs. lexical phonology in the current experiment, in the lexical phonology condition from Experiment 4, and the no pre-exposure condition provided by Day 1 of Experiment 3. Performance was better in all three pre-exposure conditions than in the no pre-exposure condition, particularly early in training. There is some suggestion that semantic pre-exposure was more beneficial than lexical phonology pre-exposure, both within the current experiment and when compared to Experiment 4.

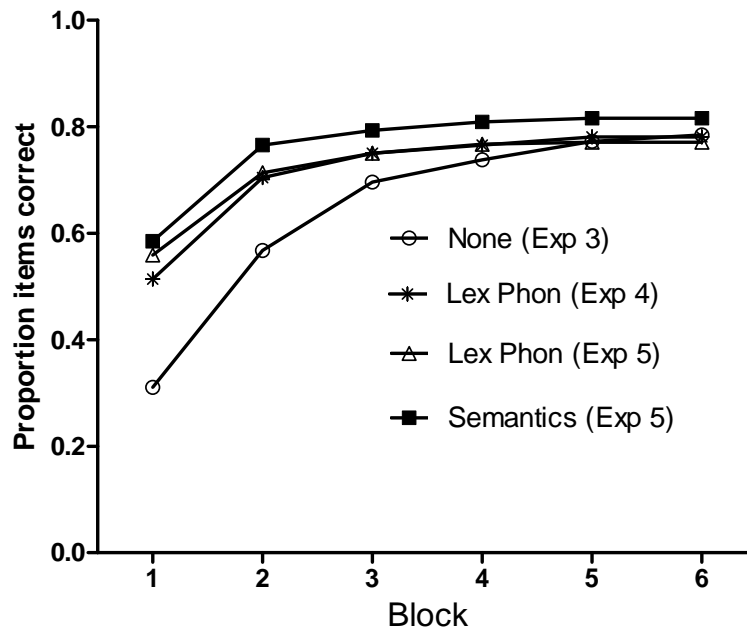


Figure 6.2. Accuracy during training in the current experiment, lexical phonology pre-exposure from Experiment 4, and no pre-exposure from Experiment 3.

Three analyses of variance were conducted to examine the effect of pre-exposure over the course of training (Block 1 – Block 6). The first analysis focused on data from the current experiment and examined the effect of the within subject semantic vs. lexical phonology pre-exposure manipulation. In this analysis, pre-exposure

and block were treated as within subject factors in both subjects and items analysis. Note that pre-exposure was within item and within subject because 16 participants received half the items in the semantic condition and half in the lexical phonology condition, and the other 16 participants received the reverse assignment. This analysis demonstrated that accuracy was higher in the semantic condition than in the lexical phonology condition, $F_p(1, 31) = 5.19, p < .05, \eta_p^2 = .14, F_i(1, 71) = 3.91, p = .05, \eta_p^2 = .05$. Accuracy improved between Blocks 1 and 3 at which point it reached asymptote, $F_p(5, 155) = 27.44, p < .001, \eta_p^2 = .47, F_i(5, 355) = 142.43, p < .001, \eta_p^2 = .67$. There was no interaction between block and pre-exposure, $F_s < 1$.

The second analysis compared performance in the semantic condition from the current experiment with performance in the lexical phonology condition from Experiment 4 and the no pre-exposure condition provided by Day 1 of Experiment 3. In this second analysis, pre-exposure was a between subjects factor in the subjects analysis and a within item factor in the items analysis; block was a repeated measures factor in both analyses. Note that 16 individuals participated in the lexical phonology condition from Experiment 4, and the no pre-exposure condition from Experiment 3, whereas 32 individuals participated in the current experiment. However, only half the items contributed to the semantic condition in the current experiment (with the other items contributing to the lexical phonology condition). Therefore, the overall power was similar across pre-exposure conditions.

The main effect of pre-exposure was significant, $F_p(2, 61) = 8.08, p < .001, \eta_p^2 = .21, F_i(2, 142) = 17.67, p < .001, \eta_p^2 = .20$. Pairwise comparisons demonstrated that performance in the semantic condition was enhanced relative to the lexical phonology condition from Experiment 4, and that both conditions outperformed the no pre-exposure condition. The main effect of block was significant, $F_p(1.69, 103.34) = 87.84, p < .001, \eta_p^2 = .59, F_i(1.89, 134.17) = 251.44, p < .001, \eta_p^2 = .78$, as was the interaction between block and pre-exposure, $F_p(10, 305) = 5.15, p < .001, \eta_p^2 = .15, F_i(10, 710) = 22.24, p < .001, \eta_p^2 = .24$. Pairwise comparisons demonstrated that in the first two training blocks, accuracy was higher in the semantic condition than in the lexical phonology condition, which in turn was higher than the no pre-exposure condition. However, in Blocks 3 and 4, accuracy in the semantic condition was higher than in the no pre-exposure condition, with accuracy in the lexical phonology condition falling in between. The effect of pre-exposure was not significant in Blocks 5 or 6.

A third analysis compared performance in the lexical phonology condition from the current experiment with the lexical phonology condition from Experiment 4, again with reference to the no pre-exposure condition provided by Day 1 of Experiment 3. Performance in the two lexical phonology conditions did not differ, but both outperformed the no pre-exposure condition, $F_p(2, 61) = 3.26, p = .05, \eta_p^2 = .10, F_i(2, 142) = 9.28, p < .001, \eta_p^2 = .12$. The main effect of block was significant, $F_p(1.66, 101.21) = 94.68, p < .001, \eta_p^2 = .61, F_i(1.91, 135.79) = 245.23, p < .001, \eta_p^2 = .78$, as was the

interaction between block and pre-exposure, $F_p(10, 305) = 6.27, p < .001, \eta_p^2 = .17, F_i(10, 710) = 22.32, p < .001, \eta_p^2 = .24$. Pairwise comparisons demonstrated that the main effect of pre-exposure was significant in Blocks 1 and 2 but not from Block 3 onwards.

To summarise, semantic pre-exposure in the form of definitions provided greater benefit than lexical phonology pre-exposure in the early and middle stages of training. Performance following lexical phonology pre-exposure was also enhanced relative to no pre-exposure, but this was less pronounced by the middle of training. By the end of training, performance was equivalent in pre-exposure and no pre-exposure conditions.

Effects of pre-exposure, frequency, and consistency at the beginning of training.

A similar set of three analyses was conducted to examine the effect of pre-exposure on items containing vowels of differing frequency and consistency in the first block of training. Frequency and consistency were treated as within subject factors in the subjects analyses and between items factors in the items analyses. Relevant data from all three analyses are contained within Figure 6.3. Examination of this figure suggests that both semantic and lexical phonology pre-exposure in the current experiment, and lexical phonology pre-exposure in Experiment 4, increased reading accuracy in Block 1, relative to the no pre-exposure condition provided by Day 1 of Experiment 3. This facilitative effect of pre-exposure appears somewhat more pronounced in the semantic than lexical phonology conditions, but does not differ according to item frequency or consistency.

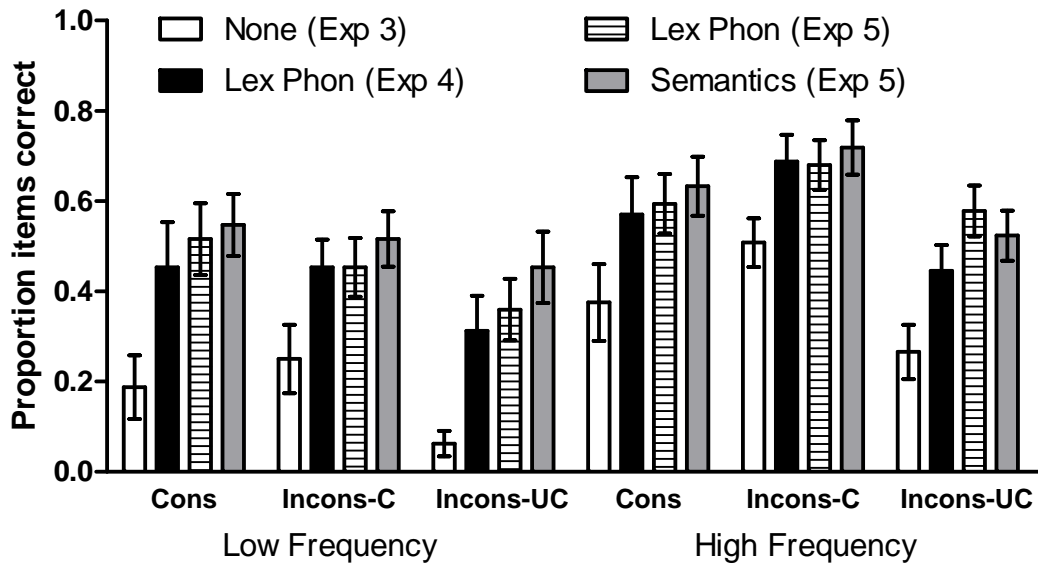


Figure 6.3. Accuracy at the beginning of training ($\pm SE$) in the current experiment, lexical phonology pre-exposure from Experiment 4, and no pre-exposure from Experiment 3, as a function of frequency and consistency.

In the first analysis, which considered the effect of the within subject semantic vs. lexical phonology manipulation, the main effect of pre-exposure was not significant, $F_p(1, 31) = 1.73$, $p = .19$, $\eta_p^2 = .05$, $F_i(1, 66) = 1.01$, ns . Pre-exposure did not interact with frequency, $F_s < 1$, or consistency, $F_s < 1$, and the three-way interaction between these variables was also non-significant, $F_s < 1$.

Turning to the second analysis, performance was equivalent in the semantic condition from the current experiment and the lexical phonology condition from Experiment 4; both improved performance relative to no pre-exposure, $F_p(2, 61) = 9.34$, $p < .001$, $\eta_p^2 = .23$, $F_i(2, 132) = 49.38$, $p < .001$, $\eta_p^2 = .43$. Pre-exposure did not interact with frequency, $F_p(2, 61) = 1.16$, ns , $F_i(2, 132) = 1.28$, ns , or consistency, $F_s < 1$, and the three-way interaction between these variables was non-significant, $F_s < 1$.

In the third analysis, performance in the two lexical phonology conditions did not differ and both provided significant benefit relative to no pre-exposure, $F_p(2, 61) = 7.03$, $p < .01$, $\eta_p^2 = .19$, $F_i(2, 132) = 47.68$, $p < .001$, $\eta_p^2 = .42$. The effect of pre-exposure did not interact with frequency, $F_s < 1$, or consistency, $F_s < 1$, and the three-way interaction between these variables was also non-significant, $F_s < 1$.

In all three analyses, accuracy was higher for items containing high frequency vowels than for those containing low frequency vowels. Performance was also better on consistent and inconsistent-conditioned items than on inconsistent-unconditioned items. The two-way interaction between frequency and consistency was not significant in any of these analyses. The subjects and items statistics for these frequency and consistency effects are summarised in Appendix 6.B, Table 6.B1.

To summarise, in the first block of training, semantic pre-exposure provided equivalent benefit to lexical phonology pre-exposure, both within the current experiment and when compared to Experiment 4. All pre-exposure conditions increased reading accuracy across item types, relative to no pre-exposure.

Effects of pre-exposure, frequency and consistency at the end of training. In Figure 6.4 it can be seen that semantic pre-exposure in the current experiment and lexical phonology pre-exposure in Experiment 4 seem to have enhanced end of training performance on items containing low frequency inconsistent-unconditioned vowels, relative to the no pre-exposure baseline. In contrast, lexical phonology pre-exposure in the current experiment seems to have provided less benefit for such items.

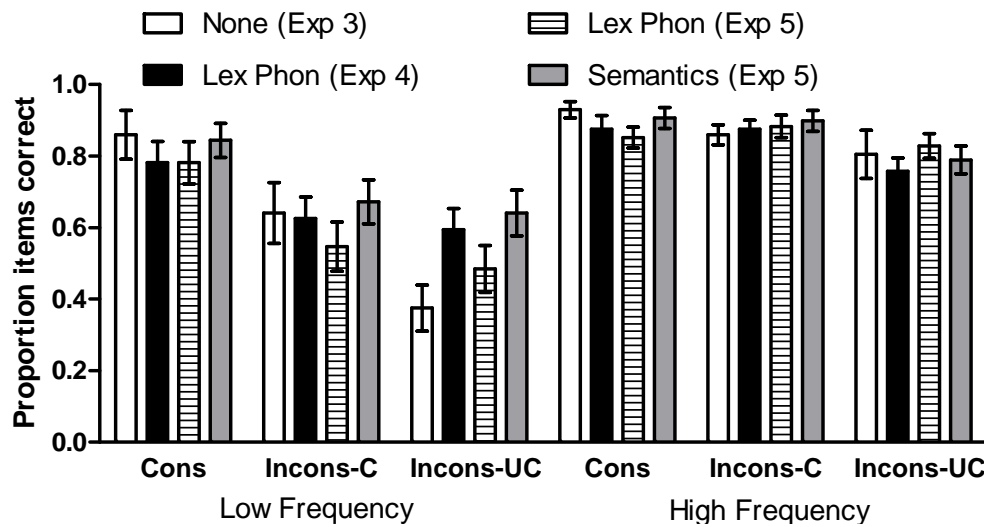


Figure 6.4. Accuracy at the end of training (\pm SE) in the current experiment, lexical phonology pre-exposure from Experiment 4, and no pre-exposure from Experiment 3, as a function of frequency and consistency.

As in preceding sections, three sets of analyses were conducted to examine the effect of pre-exposure on performance in the block at which 70% accuracy was achieved. The first analysis demonstrated that in the current experiment, accuracy was higher in the semantic condition than in the lexical phonology condition, $F_p(1, 31) = 6.76$, $p = .01$, $\eta_p^2 = .18$, $F_i(1, 66) = 7.73$, $p < .01$, $\eta_p^2 = .11$. The effect of pre-exposure interacted with frequency, $F_p(1, 31) = 5.38$, $p < .05$, $\eta_p^2 = .15$, $F_i(1, 66) = 5.37$, $p < .05$, $\eta_p^2 = .08$, and was only significant in items containing low and not high frequency vowels. Pre-exposure did not interact with consistency, $F_s < 1$. Although the three-way interaction between pre-exposure, frequency, and consistency was not significant, $F_p < 1$, $F_i(2, 66) = 1.45$, *ns*, pairwise comparisons indicated that the advantage of the semantic over the lexical phonology condition in items containing low frequency vowels was only significant for inconsistent items. This can be seen clearly in Figure 6.4.

In the second analysis, neither the main effect of pre-exposure (semantics Exp 5 vs. lexical phonology Exp 4 vs. none Exp 3), $F_p(2, 61) = 2.20, p = .12, \eta_p^2 = .07, F_i(2, 132) = 2.26, p = .11, \eta_p^2 = .03$, nor the interaction between pre-exposure and frequency, $F_p(2, 61) = 1.19, ns, F_i(2, 132) = 2.09, p = .13, \eta_p^2 = .03$, were significant. The interaction between pre-exposure and consistency was marginal by-items, $F_i(4, 132) = 2.18, p = .08, \eta_p^2 = .06$, but was not significant by-subjects, $F_p(4, 122) = 1.05, ns$. The three-way interaction between pre-exposure, frequency, and consistency was significant by-items, $F_i(4, 132) = 2.68, p < .05, \eta_p^2 = .08$, but not by-subjects, $F_p(4, 122) = 1.21, ns$. Pairwise comparisons across items showed there to be no difference between semantic pre-exposure in the current experiment and lexical phonology pre-exposure in Experiment 4, but that both benefited performance on items containing low frequency inconsistent-unconditioned vowels relative to no pre-exposure.

In the third analysis, the main effect of pre-exposure (lexical phonology Exp 5 vs. lexical phonology Exp 4 vs. none Exp 3) was not significant, $F_s < 1$. There was no interaction between pre-exposure and frequency, $F_p < 1, F_i(2, 132) = 1.76, p = .18, \eta_p^2 = .03$. The interaction between pre-exposure and consistency was marginal by-items, $F_i(4, 132) = 2.33, p = .06, \eta_p^2 = .07$, but non-significant by-subjects, $F_p < 1$. The interaction between pre-exposure, frequency, and consistency was significant by-items, $F_i(4, 132) = 2.50, p < .05, \eta_p^2 = .07$, but not by-subjects, $F_p < 1$. Pairwise comparisons across items demonstrated that, unlike in Experiment 4, pre-exposure to lexical phonology in the

current experiment did not benefit reading accuracy for items containing low frequency inconsistent-unconditioned vowels at the end of training.

In all three analyses, performance was better for items containing high frequency vowels than for those containing low frequency vowels. Performance was also better for consistent items than inconsistent-conditioned items, which in turn outperformed inconsistent-unconditioned items. The interaction between frequency and consistency was also significant across analyses. Pairwise comparisons demonstrated that the frequency effect was only significant within inconsistent items. In addition, within items containing low frequency vowels, accuracy was higher for consistent items than for both types of inconsistent item, whereas within items containing high frequency vowels, accuracy was higher for consistent items than for inconsistent-unconditioned items, but inconsistent-conditioned items fell in between. Subjects and items statistics for these frequency and consistency effects are summarised in Appendix 6.B, Table 6.B2.

To summarise, in the current within subject experiment, pre-exposure to semantics was more beneficial than pre-exposure to lexical phonology at the end of training. This benefit was particularly clear for items containing low frequency inconsistent vowels. However, when compared between subjects to the lexical phonology condition from Experiment 4, no benefit of semantic pre-exposure was evident. Furthermore, while semantic pre-exposure in the current experiment and lexical phonology pre-exposure in Experiment 4 enhanced performance on items containing low frequency inconsistent-unconditioned vowels (relative to the no pre-

exposure condition), this was not the case following lexical phonology pre-exposure in the current experiment. These results suggest that within the current experiment, the end of training advantage of semantic over lexical phonology pre-exposure for items containing low frequency inconsistent vowels was not a genuine semantic effect. Instead, it resulted from the within subject manipulation reducing the benefit of lexical phonology pre-exposure. When attention is focused during lexical phonology pre-exposure (as in Experiment 4), it provides equivalent benefit to semantic pre-exposure for performance at the end of training. This benefit is specific to items containing low frequency inconsistent-unconditioned vowels.

Old-new decision

Trials with RTs more than two standard deviations away from each participant's mean were excluded from the analysis. In the current experiment this constituted 4.6% of trials. One-sample t-tests confirmed that accuracy in the current experiment was above chance on both trained, $t_p(31) = 18.47$, $p < .001$, and untrained items, $t_p(31) = 8.84$, $p < .001$. Paired sample t-tests showed that accuracy for trained, $M = .80$ (.09), and untrained items, $M = .76$ (.17), did not differ, $t_p(31) = 1.15$, ns. However, RTs were faster to trained, $M = 3483$ (851), than untrained items, $M = 4398$ (1399), $t_p(31) = 6.06$, $p < .001$.

As with the training data, three sets of analyses were conducted to examine the effect of pre-exposure on old-new decision accuracy and mean RTs to correct trained items. In the first analysis, which examined the within subject pre-exposure manipulation in the current experiment, frequency and consistency were included as

factors in the analysis. This was in order to assess whether the frequency and consistency effects seen in old-new decision in Experiment 4 replicated and were strengthened by the inclusion of a greater number of items. Several participants in the current experiment made no correct responses to a particular item type. This resulted in several missing data points in the RT subjects analysis (5.5%) which were replaced with the mean RT for that item type and pre-exposure condition. The results of this first analysis are summarised in Figure 6.5.

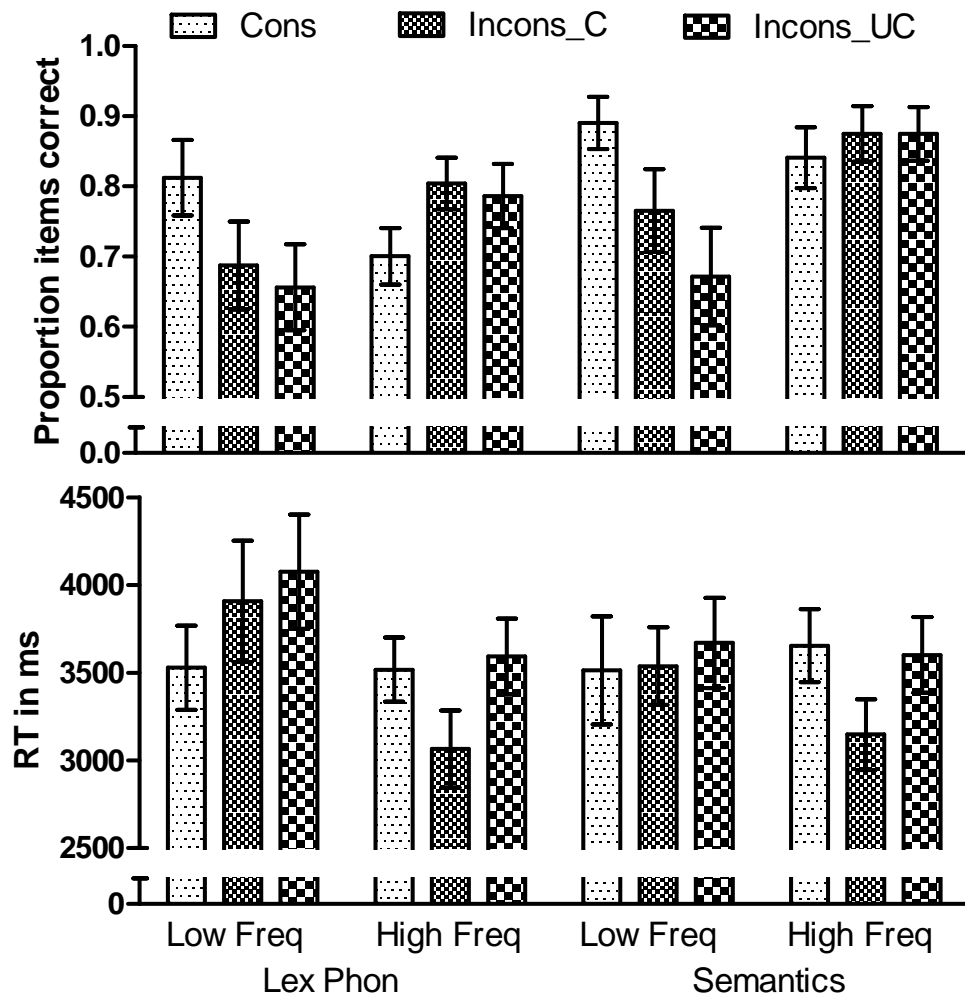


Figure 6.5. Old-new decision accuracy and RTs (\pm SE) in the current experiment as a function of pre-exposure (within subject), frequency, and consistency.

Accuracy was higher in the semantic than the lexical phonology condition, $F_p(1, 31) = 10.88, p < .01, \eta_p^2 = .26, F_i(1, 66) = 10.32, p < .01, \eta_p^2 = .14$. The effect of pre-exposure on accuracy did not interact with frequency or consistency and there was no three-way interaction between these variables (all $F_s < 1$). The main effect of pre-exposure was not reflected in the latency data, $F_p(1, 31) = 1.08, ns, F_i < 1$. In this latency analysis, the interaction between pre-exposure and frequency was not significant by-subjects, $F_p(1, 31) = 2.18, p = .15, \eta_p^2 = .07$, but was marginal by-items, $F_i(1, 66) = 3.19, p = .08, \eta_p^2 = .05$. Pairwise comparisons across items demonstrated that RTs to items containing low frequency vowels were marginally faster in the semantic condition than in the lexical phonology condition, but RTs to items containing high frequency vowels were unaffected by pre-exposure. In addition, in the lexical phonology condition, RTs were faster to items containing high frequency vowels than to items containing low frequency vowels, but in the semantic condition, RTs were unaffected by frequency. Neither the interaction between pre-exposure and consistency, nor the three-way interaction between pre-exposure, frequency, and consistency, was significant in the latency analysis (all $F_s < 1$).

Considering the overall effects of frequency and consistency, accuracy was higher, $F_p(1, 31) = 5.42, p < .05, \eta_p^2 = .15, F_i(1, 66) = 4.36, p < .05, \eta_p^2 = .06$, and RTs were faster, $F_p(1, 31) = 9.15, p < .01, \eta_p^2 = .23, F_i(1, 66) = 4.21, p < .05, \eta_p^2 = .06$, for items containing high frequency vowels than for those containing low frequency vowels. The main effect of consistency was not significant in either the accuracy, $F_p(2, 62) = 1.49, ns$,

$F_i(2, 66) = 1.04$, ns, or RT analysis, $F_p(2, 62) = 1.87$, $p = .16$, $\eta_p^2 = .06$, $F_i(2, 66) = 2.20$, $p = .12$, $\eta_p^2 = .06$. In the accuracy analysis, there was a significant interaction between frequency and consistency, $F_p(2, 62) = 6.92$, $p < .01$, $\eta_p^2 = .18$, $F_i(2, 66) = 5.12$, $p < .01$, $\eta_p^2 = .14$. Pairwise comparisons and examination of effect sizes demonstrated that the facilitative effect of frequency was significant across items types, but was most pronounced in inconsistent-unconditioned items, followed by inconsistent-conditioned items, and was least strong in consistent items. There was no effect of consistency within items containing high frequency vowels, but within items containing low frequency vowels, accuracy was highest for consistent items, followed by inconsistent-conditioned items, and lowest for inconsistent-unconditioned items.

In the RT analysis, the interaction between frequency and consistency was not significant by-subjects, $F_p(2, 62) = 2.20$, $p = .12$, $\eta_p^2 = .07$, but was by-items, $F_i(2, 66) = 4.25$, $p < .05$, $\eta_p^2 = .11$. Pairwise comparisons across items indicated that the facilitative effect of frequency was only significant within inconsistent-conditioned items. There was no effect of consistency within items containing low frequency vowels, but within items containing high frequency vowels, accuracy was higher for inconsistent-conditioned items than for inconsistent-unconditioned and consistent items, which did not differ from each other.

The second and third analyses compared the no pre-exposure condition provided by Experiment 3, and the lexical phonology condition from Experiment 4, with the semantic and lexical phonology conditions from the current experiment,

respectively. The results of these analyses are summarised in Figure 6.6. In these comparisons, data were collapsed across frequency and consistency. This was intended to clarify any pre-exposure effects which might have been dampened by the fact that fewer items were included in old-new decision in the no pre-exposure and lexical phonology conditions from Experiments 3 and 4 (trained items $n = 12$) than in current experiment (trained items $n = 36$). This difference between item numbers also meant that items analyses could not be conducted.

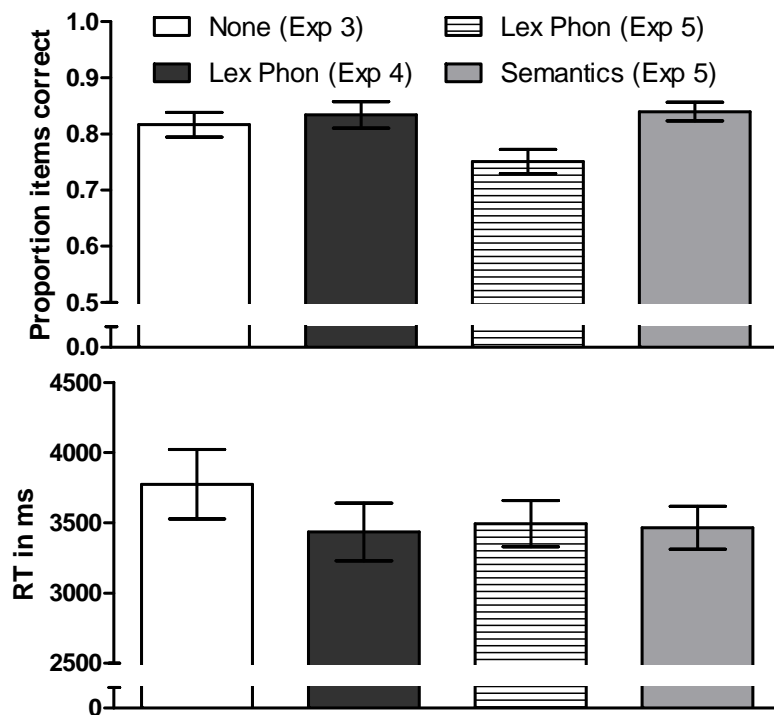


Figure 6.6. Old-new decision accuracy and RTs (\pm SE) in the current experiment (Exp 5), lexical phonology pre-exposure from Experiment 4, and no pre-exposure from Experiment 3.

In the second analysis there was no difference between the semantic condition from the current experiment, the lexical phonology condition from Experiment 4, and

the no pre-exposure condition (Day 1 of Experiment 3) in either accuracy, $F_p < 1$, or RTs, $F_p < 1$. In the third analysis, accuracy was lower in the lexical phonology condition from the current experiment than in the equivalent between subjects condition from Experiment 4, with accuracy in the no pre-exposure condition falling in between, $F_p(2, 61) = 3.91, p < .05, \eta_p^2 = .11$. RTs did not differ between these three conditions, $F_p < 1$.

To summarise, in the current experiment, old-new decision accuracy was higher in the semantic condition than in the lexical phonology condition. Within items containing low frequency vowels, RTs were also faster in the semantic condition, although this effect was marginal across items and unreliable across subjects. When compared to the lexical phonology condition from Experiment 4, semantic pre-exposure did not benefit discrimination accuracy or RTs. Furthermore, accuracy was lower in the lexical phonology condition in the current experiment than in the equivalent condition from Experiment 4. These findings support those from the analyses of reading accuracy at the end of training, and suggest that the advantage of the semantic over the lexical phonology condition in the current experiment was not a genuine semantic effect. Instead, it resulted from depressed performance in the lexical phonology condition.

Generalization

As the lexical phonology vs. semantics manipulation in this experiment was within subject, its effect on generalization could not be assessed. Instead, an analysis compared generalization in the current experiment with the no pre-exposure baseline (Day 1 - Experiment 3). The results of this analysis are summarised in Figure 6.7. Generalization was marginally lower in the current experiment than on Day 1 of

Experiment 3, $F_p(1, 46) = 3.65$, $p = .06$, $\eta_p^2 = .07$, $F_i(1, 18) = 3.50$, $p = .08$, $\eta_p^2 = .18$. This effect did not interact with frequency, $F_s < 1$, or consistency, $F_p(2, 92) = 1.13$, ns , $F_i(2, 18) = 1.14$, ns . However, the three-way interaction between these variables was significant by-subjects, $F_p(2, 92) = 3.18$, $p < .05$, $\eta_p^2 = .07$, although not by-items, $F_i(2, 18) = 1.82$, $p = .19$, $\eta_p^2 = .17$. Pairwise comparisons across subjects indicated that the lower accuracy in the current experiment was only significant in items containing low frequency inconsistent vowels. This can be seen clearly in Figure 6.7.

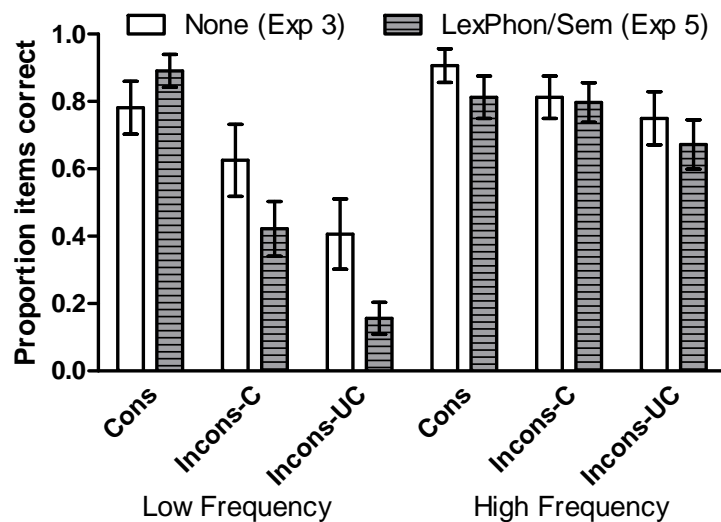


Figure 6.7. Generalization accuracy ($\pm SE$) in the current experiment and the no pre-exposure condition from Experiment 3, as a function of frequency and consistency.

Accuracy was higher for generalization items containing high frequency vowels than for those containing low frequency vowels, $F_p(1, 46) = 27.68$, $p < .001$, $\eta_p^2 = .38$, $F_i(1, 18) = 77.06$, $p < .001$, $\eta_p^2 = .81$. The main effect of consistency was also significant, $F_p(1.28, 79.42) = 18.13$, $p < .001$, $\eta_p^2 = .28$, $F_i(2, 18) = 53.02$, $p < .001$, $\eta_p^2 = .86$, and arose

from the fact that accuracy was higher for items containing consistent vowels than for those containing inconsistent-conditioned vowels, which in turn outperformed items containing inconsistent-unconditioned vowels. The main effects of frequency and consistency were qualified by a significant interaction, $F_p(2, 92) = 10.99, p < .001, \eta_p^2 = .19, F_i(2, 18) = 18.11, p < .001, \eta_p^2 = .67$. Pairwise comparisons indicated that the frequency effect was only significant in items containing inconsistent vowels and that the consistency effect was only significant in items containing low frequency vowels.

Discussion

The aim of this experiment was to examine the role of semantic vs. phonological familiarity in orthographic learning. Natural language research suggests a role for semantics in reading aloud (Frost et al., 2005; McKay et al., 2007; Patterson et al., 2006; Strain et al., 1995;2002; Woollams et al., 2007) and visual word recognition (Balota et al., 2004; Cortese & Khanna, 2007; Pexman et al., 2008). However, such research is compromised by problems in measuring and controlling for the many variables that influence these skills. In particular, research has left open the possibility that it is the strength of phonological rather than semantic representations that supports orthographic processing. Using a novel word learning paradigm, McKay et al. (2008) found that participants learned to read inconsistent novel words more successfully if they had previously learned their meanings (definitions) rather than just their sounds (repetition of phonological forms). Discrimination of trained from untrained items was also faster and more accurate when trained items had been given a meaning. This

provided the first evidence that semantic familiarity supports orthographic processing over and above phonological familiarity.

Experiment 4 failed to replicate this effect using the artificial orthography paradigm. Participants who were pre-exposed to lexical phonology showed the same improvement to subsequent orthographic processing as participants pre-exposed to semantics, relative to a no pre-exposure baseline. It is important to remember that this was not a null effect: both pre-exposure conditions supported learning to read aloud, and by the end of training facilitation was specific to items containing low frequency inconsistent vowels. Pre-exposure also increased the speed with which participants could discriminate trained from untrained items. Thus, in Experiment 4, phonological familiarity conferred exactly those benefits that natural language research and McKay et al. (2008) have reported for semantic familiarity.

The current experiment explored reasons for the discrepancy between the results of Experiment 4 and those reported by McKay and colleagues (2008). As discussed earlier, one possibility is that Experiment 4 failed to find an additional effect of semantics, over and above that conferred by phonological familiarity, because the novel objects used to represent word meanings did not induce strong enough semantic representations. However, a second possibility is that McKay et al.'s within subject pre-exposure manipulation reduced the benefit of the lexical phonology condition. On this view, the benefit of definitions pre-exposure arose not from their semantic content but because learning them focused attention on semantic items and reduced attention to lexical phonology items. To explore these possibilities, the current experiment used a

definitions task to instantiate semantics at pre-exposure, rather than the object-referent task used in Experiment 4. This was compared to lexical phonology pre-exposure, using a within subject design, rather than the between subjects design used in Experiment 4. To assess reasons for any effects of the within subject pre-exposure manipulation, comparisons were also made between subjects, using data from the semantic and lexical phonology conditions from the current experiment, and no pre-exposure and lexical phonology conditions from Experiments 3 and 4.

The following discussion first explores whether semantic or phonological familiarity supports learning to read aloud. The influence of semantic/phonological pre-exposure early and late in training is then addressed. Discrimination performance is next examined with respect to the influences of pre-exposure, frequency, and consistency. Finally, generalization is considered.

The effects of phonological vs. semantic familiarity on learning to read aloud

In the current experiment, semantic pre-exposure provided greater benefit for reading accuracy during training than lexical phonology pre-exposure. In the block in which 70% accuracy was achieved, the facilitative effect of semantics was specific to items containing low frequency inconsistent vowels. Taken alone, these findings mirror those of McKay et al. (2008) and support the idea that semantics plays a role in reading aloud, over and above any benefits conferred through phonological familiarity. Moreover, they suggest that at higher levels of proficiency, this semantic support is particularly important for items with low frequency inconsistent spelling-sound

mappings, as suggested by the triangle model (Harm & Seidenberg, 2004; Plaut et al., 1996).

However, as discussed earlier, the facilitative effect of semantics observed in the current experiment could have arisen for two reasons. It could have been a genuine semantic effect arising from the strong semantic representations created by the novel definitions. However, it could also have been a result of the within subject pre-exposure manipulation, in which the demanding definition learning task focused attention on semantic items but reduced attention to lexical phonology items.

In support of the first suggestion, an analysis of performance over the entire orthography training phase demonstrated that the semantic condition also outperformed the lexical phonology condition from Experiment 4, in both early and middle training blocks. This supports the idea that familiarity with word meanings plays a role in learning to read over and above familiarity with word sounds. However, analyses of performance at the beginning and end of training complicate the picture. In both Block 1 and the block in which 70% accuracy was achieved, semantic pre-exposure in the current experiment conferred equivalent benefit to lexical phonology pre-exposure in Experiment 4, relative to the no pre-exposure condition. Furthermore, although lexical phonology pre-exposure facilitated end of training performance on items containing low frequency inconsistent-unconditioned vowels in Experiment 4, this was not the case for lexical phonology pre-exposure in the current experiment. The end of training results therefore support the second suggestion, that the within subject pre-exposure manipulation reduced the benefit of lexical phonology pre-exposure. Thus, the

benefit that semantic pre-exposure conferred to end of training performance on items containing low frequency inconsistent vowels was not a genuine semantic effect. Instead, it was an artefact induced by the fact that learning definitions concentrated attention on semantic items and decreased attention to lexical phonology items.

Overall then, when compared to Experiments 3 and 4, the results of Experiment 5 provide support for both views. Semantic pre-exposure benefited learning to read aloud more than lexical phonology pre-exposure (in either the current experiment or Experiment 4) when performance was analysed over the entire orthography training phase. But, when considering end of training performance on items containing low frequency inconsistent vowels, semantic and lexical phonology pre-exposure conferred equivalent benefit, relative to the no pre-exposure baseline. This finding is particularly interesting given that previous research has suggested a role for semantics in skilled reading of low frequency inconsistent words.

Why were specific semantic effects absent when considering end of training performance on items containing low frequency inconsistent vowels? One possibility is that higher proficiency on the artificial orthography might be required for these specific semantic effects to outweigh those conveyed by phonological familiarity. Some support for this suggestion comes from a more in depth consideration of performance over the course of training. Early in training, accuracy was higher in the semantic condition than in the lexical phonology condition from Experiment 4, which in turn outperformed the no familiarity condition. At this point, the advantage of semantic over lexical phonology pre-exposure was less pronounced than the advantage of lexical phonology over no pre-

exposure. However, in the middle stages of training, only the advantage of the semantic over the no pre-exposure condition persisted, with performance in the lexical phonology condition falling in between. Thus, the benefit of semantic relative to phonological familiarity increased over the course of training, but was not significant in final training blocks. Perhaps then with extended training, the benefit conveyed by semantic familiarity to end of training performance on items containing low frequency inconsistent vowels would outweigh that conferred by phonological familiarity. Moreover, increasing proficiency might make it possible to assess pre-exposure effects in terms of both reading accuracy and speed. At the relatively low proficiency levels examined thus far, reading aloud is too laboured to consider response latencies.

Changes in the effect of pre-exposure during training

As in Experiment 4, pre-exposure improved performance across items at the beginning of training, but by the end of training, facilitation was restricted to items containing low frequency inconsistent vowels. These findings suggest that in the early stages of learning, phonological/semantic familiarity benefits reading of all items. The restriction of effects to items with inconsistent and low frequency spelling-sound mappings emerges as learning progresses. As discussed in Chapter 5, this possibility has not been explored in previous research which has largely focused on fairly competent readers. These results remind us that the effect of semantics, and the distinction between consistent and inconsistent words, should not be thought of as all-or-nothing. They support the idea, embodied in the triangle model framework, that all words rely on the same sets of representations and that the "... division of labour..." (Plaut et al., 1996,

p. 58) between such representations (phonological, orthographic, semantic) may change over time. However, the present findings also challenge Plaut et al.'s assumption that semantic knowledge may only make a substantial contribution to reading aloud once orthography-phonology mappings have developed to some degree. Instead they suggest that the system uses all knowledge available wherever it is most efficient to do so.

Discrimination

In the current experiment, old-new decision accuracy was higher following semantic pre-exposure than lexical phonology pre-exposure. RTs were also faster in the semantic condition, although this was only a marginal effect and was limited to items containing low frequency vowels. These findings are similar to those of McKay et al. (2008) and taken alone, they suggest that familiarity with word meanings supports word recognition. However, as with reading accuracy at the end of training, when compared to the lexical phonology condition from Experiment 4, no semantic advantage was observed in either the accuracy or latency data. Furthermore, accuracy in the lexical phonology condition was lower in the current experiment than in the equivalent condition from Experiment 4. Overall, these results suggest that in the current experiment, rather than definitions enhancing discrimination due to their semantic content, including definitions served to depress performance on items in the lexical phonology condition.

In Experiment 4, lexical phonology pre-exposure did not affect old-new decision accuracy but did decrease RTs. In the analyses reported in the current chapter, a comparison between the semantic condition from the current experiment, the lexical

phonology condition from Experiment 4, and the no pre-exposure condition, revealed no significant effects of pre-exposure on discrimination accuracy or RTs. Thus, when data are considered across experiments, they do not provide convincing evidence that either phonological or semantic familiarity benefit recognition of the orthographic forms of trained items. Reasons for this will be considered later in this section.

Turning to the effects of frequency and consistency in old-new decision, it is helpful to first summarise the effects seen in earlier experiments. On Day 1 of Experiment 3, accuracy was unaffected by frequency or consistency, but RTs were faster to items containing high frequency than low frequency vowels. With the increased power provided by more participants in Experiment 4, a frequency by consistency interaction emerged in the accuracy data: within items containing low frequency vowels, accuracy was higher for consistent than inconsistent items. The facilitative effect of frequency on RTs observed in Experiment 3 was also replicated.

The inclusion of more items in the current experiment served to strengthen the effects of frequency and consistency on accuracy, relative to those observed in Experiments 3 and 4. Accuracy was higher for consistent than inconsistent items (as in Experiment 4) but further differentiation also emerged, with inconsistent-conditioned items outperforming inconsistent-unconditioned items. In addition, in the current experiment the effect of consistency was restricted to items containing low frequency vowels. Furthermore, accuracy was higher for items containing high frequency vowels, and this effect was most pronounced in inconsistent-unconditioned items. These effects mirror those observed in typical lexical decision (Lacruz & Folk, 2004).

Turning to the latency data, RTs were faster to items containing high frequency vowels, as in Experiments 3 and 4. However, in the current experiment this effect was restricted to inconsistent-conditioned items. Consistency effects were also evident, but these were restricted to items containing high frequency vowels, in which inconsistent-conditioned items outperformed both consistent and inconsistent-unconditioned items. Thus, although old-new decision RTs were influenced by frequency and consistency, this was not in entirely the same way as seen in lexical decision, where latency and accuracy data pattern together (Lacruz & Folk, 2004). This may be due to the restricted set of items that comprised the artificial orthography. In particular, it seems that the high salience of the repeated consonant-vowel onset that characterised inconsistent-conditioned items increased the speed with which participants could recognise such items. A reconsideration of consistency effects during training also suggests that this effect might have been exaggerated by the difficulty of old-new decision. As discussed in detail in Experiment 2, early in training, items containing high frequency inconsistent-conditioned vowels outperformed items containing high frequency consistent vowels, but by the end of training this advantage had dissipated. This suggests that the high salience of inconsistent-conditioned vowels is most beneficial at lower levels of proficiency. Thus, if old-new decision proficiency increased, we might expect RTs to be fastest to consistent, rather than inconsistent-conditioned, items.

Although old-new decision required knowledge of which items constituted “words” in the language, item recognition was not automatic, as it is in typical lexical decision. Instead, it seems that items were decoded and consciously evaluated for

orthographic and phonological familiarity. This might be key to explaining why frequency and consistency did influence performance, but phonological/semantic familiarity did not. Balota et al. (2004) conducted a large scale regression analysis to examine the predictors of naming and lexical decision in English orthography. They suggested (and found) that "... effects of spelling-to-sound consistency should be greater in naming than in lexical decision because naming requires the use of phonological information, whereas the LDT does not place the same premium on this information..."(p. 285). Furthermore, Balota et al. also stated (and again found) that, "... larger effects of these semantic variables in lexical decision performance were expected because this task places a greater emphasis on the meaningfulness of the stimulus..." (p. 311). However, in the artificial orthography it seemed necessary for participants to decode items before they could discriminate them, thus old-new decision relied heavily on phonological information. This reliance on phonological processing may explain why frequency and consistency influenced old-new decision, but semantic/phonological familiarity did not. This view predicts that old-new decision should show decreased sensitivity to frequency and consistency, and increased sensitivity to semantic/phonological familiarity with extended orthography training. As discussed in Chapter 4, developmental data support the suggestion that frequency and consistency effects in lexical decision are strongest at lower levels of proficiency (Unsworth & Pexman, 2003; Waters et al., 1984). Unfortunately, however, research has not yet considered developmental differences in semantic influences on visual word recognition.

Generalization

As in Experiment 4, generalization was poorer in the current experiment than in the no pre-exposure condition provided by Day 1 of Experiment 3. In Experiment 4, it was suggested that pre-exposure to item sounds might increase the extent to which participants read trained items by recalling that they were a valid word in the language. This might reduce the necessity to extract information about the influence of consonants on vowel pronunciations. In turn, this would be detrimental for generalization. This proposal is supported by the results of the current experiment in which decreased generalization (induced by pre-exposure) was specific to items containing low frequency inconsistent vowels. Such items could only be read correctly if participants had developed sensitivity to the influence of consonant onsets on vowel character pronunciations. It should be noted that here, as in previous experiments, only one participant explicitly deduced one of the inconsistent vowel character-sound mappings.

As discussed in Chapter 5, research on orthographic learning in children supports the idea that decreased attention to orthography-phonology mappings is detrimental for generalization. Landi et al. (2006) found that children's reading of new words was more accurate in context than in isolation, but that retention of spelling-sound knowledge was better when words were learned in isolation. However, this research also found that learning in context had a less negative effect on retention for more advanced readers. This suggests that increased proficiency in the artificial orthography should decrease the detriment conferred to generalization by pre-exposure.

Summary and conclusions

The results of this experiment, and comparisons with Experiments 3 and 4, suggest that familiarity with word meanings may benefit learning to read aloud, over and above the benefit conferred by familiarity with word sounds. However, specific semantic effects on end of training reading accuracy for items containing low frequency inconsistent vowels were not observed. Instead, lexical phonology and semantic pre-exposure conferred the same benefit, relative to the no pre-exposure baseline. There was some evidence to suggest that extended training might be necessary for a specific effect of semantics on low frequency inconsistent item reading to emerge. Overall, the results suggest that phonological familiarity is important in learning to read and that familiarity with word meanings provides some additional benefit. Natural language experiments which have not considered the role of phonological familiarity may have overestimated the effect of semantics on orthographic processing.

The comparison of results across experiments also demonstrated that interference can occur between conditions in a within subject design. The benefit of lexical phonology pre-exposure for both trained item reading and discrimination was depressed in the current experiment relative to that observed in Experiment 4. This suggests that at least some of the differences between semantic and lexical phonology pre-exposure observed in McKay et al.'s (2008) experiment may have been a result of their within subject semantic manipulation. Future work in this area should consider this possibility when designing experiments using within and between subject manipulations.

A further important finding was that pre-exposure influenced reading accuracy across all items at the beginning of training, with effects only becoming specific to items containing low frequency inconsistent vowels by the end of training. Reasons for this were considered in Chapter 5 and will be revisited in Chapter 7. Developmental changes in the influence of phonological/semantic familiarity on reading aloud represent an important area for future modelling and behavioural research.

In the current experiment, old-new decision accuracy (although not RTs) showed predictable effects of frequency and consistency, but the task was insensitive to pre-exposure. It may be that extended training is necessary for semantic/phonological familiarity to influence recognition of newly learned items in an artificial orthography. In both the current experiment and in Experiment 4, generalization was reduced following pre-exposure, relative to no pre-exposure. Again, extended training should eliminate this effect.

CHAPTER 7: GENERAL DISCUSSION

This discussion will consider the key methodological and theoretical issues arising from the experiments reported in this thesis. Particular attention will be given to evaluating whether/how the methodologies used enabled consideration of the questions of theoretical interest. Limitations of these methods and suggestions for future research form an integral part of these discussions. Consideration is also given to how the results presented in this thesis fit with existing models of reading aloud.

Why develop an artificial orthography paradigm and was it useful?

The artificial orthography methodology was developed in response to problems identified with existing research using natural alphabetic scripts. These problems were illustrated by Experiment 1, in which adults learned to read novel words written in English orthography. The novel words were assigned consistent or inconsistent pronunciations, and participants were pre-exposed to either high or low imageability definitions, (semantics) or to item sounds (lexical phonology), or received no pre-exposure. Reading accuracy during training was higher in the semantic conditions than in the lexical phonology and no pre-exposure conditions, offering some support for the view that familiarity with word meanings supports learning to read aloud. However, some unexpected findings cast doubt on this interpretation. There was no advantage for learning high imageability definitions over low imageability definitions, suggesting that the beneficial effect of definitions pre-exposure may not have been semantic in origin.

Furthermore, pre-exposure did not have a positive influence on post-training naming or discrimination latencies.

It was suggested that some specific methodological problems might have contributed to the lack of clarity in semantic effects observed in Experiment 1. First, the binary consistency manipulation was inadequate because consistency is a graded variable that can vary in multiple ways (Jared et al., 1990). Second, novel word reading was highly accurate from the beginning of the experiment; it therefore failed to address the influence of semantics on orthographic *learning*. Third, although research suggests that semantic influences on reading aloud vary according to word frequency and spelling-sound consistency, we in fact know very little about how sensitivity to these more basic variables develops. The conclusion drawn at the end of Experiment 1 was that a more controlled paradigm is needed to investigate the influence of frequency, consistency, and semantics on orthographic learning.

Experiments 2 and 3 introduced such a method and used it to investigate the influence of spelling-sound frequency and consistency on orthographic learning and generalization. Adults learned to read novel words written in novel characters which varied in the frequency and consistency of their vowel character-sound mappings. This artificial orthography paradigm enabled complete control over exposure to the statistics of the language and direct examination of the learning process, providing clear answers to three previously open questions. First, it was demonstrated that learners could extract spelling-sound patterns through exposure to whole-word forms. Second, this spelling-sound knowledge was context-sensitive; learners detected conditional

relationships between consonant onsets and vowel pronunciations and used these when generalizing to a further set of novel words. This provides the first conclusive evidence that knowledge of context-sensitive character-sound mappings can be extracted through exposure to the orthographic form of whole words and their pronunciations. Third, the frequency and consistency of character-sound mappings interacted in their effects on learning and generalization. This demonstrates that reading aloud in the artificial orthography was subject to some of the same influences that have been observed in English (Jared, 2002; Kessler & Treiman, 2001; Weekes et al., 2006). These findings were replicated in Experiments 4 and 5, showing that the paradigm produces consistent and reliable results.

One interpretation of these findings might be that as participants were highly literate adults, they were simply mapping the artificial orthography onto their knowledge of English letter-sound mappings. This seems unlikely, especially for vowels: the vowel sounds in the artificial orthography could be represented in multiple ways in written English. For example, the phoneme /i/ can be spelled, E, EE, EA, IE, EI, EY, and /u/ can be spelled, U, O, OO, UE, OU, UI, EW, etc. Therefore, vowel character-phoneme relationships could not be mapped to English in any simple one-to-one fashion.

Another concern might be that the context-sensitive rules that governed inconsistent vowel pronunciations in the artificial orthography were deterministic. Arguably, these might be easier to abstract than the probabilistic conditional relationships that exist in English (Treiman et al., 1995). However, the less than perfect generalization observed in Experiments 2 to 5 speaks against this: although participants

did pronounce vowels according to their consonant context, they were not 100% accurate at doing so. In addition, post-experiment debriefing revealed that only 3 participants (from a total of 92 across Experiments 2 to 5) were explicitly aware of any of the consonant-vowel pronunciation rules. In sum, learning the novel orthography was achievable but not trivial, and most participants were not able to identify the rules that they had nevertheless extracted.

The lexical status of newly learned items

In experiments which examine novel word processing it is important to question whether newly learned items are treated in a similar way to existing words. If this is shown to be the case it increases the extent to which findings inform us about natural language processing. In Experiment 1, old-new decisions to trained items were faster following a masked identity prime than following a control prime. No priming effects were observed for untrained items. This mirrors effects seen in typical lexical decision in which masked identity priming decreases response latencies to words but not nonwords (Bowers, 2003; Davis et al., 2008; Forster et al., 2003; Forster, 1998; Forster & Davis, 1984). This provides some evidence that newly learned orthographic forms are lexicalised, at least when written in a familiar orthography.

Evidence for lexicalisation was less clear in Experiments 2 to 5 which used the artificial orthography. Participants learned to read the training items in a relatively short learning phase and they were then able to generalize to novel items in a test phase which followed immediately. However, in this test phase, discrimination (as measured by the old-new decision task) was laborious. This suggests that although character-

sound relationships had been extracted, whole-item orthographic representations were not strong. Furthermore, Experiment 4 demonstrated that old-new decisions in the artificial orthography were insensitive to masked identity priming. Thus, there is evidence that trained items were treated similarly to real words in Experiment 1, but not in Experiment 4 (and by inference, in Experiments 2, 3, and 5 which also used the artificial orthography).

It is possible that over time, processing of newly learned items in the artificial orthography would become more similar to that of real words. Some support for this idea is provided by previous word learning research. Gaskell and Dumay (2003) taught adults the spoken forms of novel words that overlapped closely with existing words, for example, cathedruke (cathedral) and blossail (blossom). Learning the novel words increased lexical decision latencies to the existing words a week after, but not directly following, training. These delayed interference effects suggest that it takes time for newly learned spoken words to be incorporated into the existing lexicon. Dumay and Gaskell (2007) later established that interference was dependent on a night's sleep. They suggested that overnight consolidation is necessary for new words to be integrated into the existing lexicon.

An obvious question is whether consolidation is also necessary in orthographic learning. This was explored by Bowers et al. (2005) in an experiment in which adults learned novel words written in English orthography. These novel words differed by one letter from existing English words which previously had possessed no neighbours (e.g., BANARA was trained as a novel word neighbour of BANANA, or NERDLE for NEEDLE).

Bowers et al. then tested participants' ability to semantically categorise the existing words as "natural" or "artefact" when presented with their orthographic forms. They found that learning the novel words slowed down categorization of existing words. This interference effect occurred immediately after novel word learning but was more pronounced on the following day. This supports Dumay and Gaskell's (2007) argument for overnight consolidation in new word learning. However, it also suggests that in orthographic learning some integration may take place even without overnight sleep.

Further evidence for immediate integration in orthographic learning is provided by Clay, Bowers, Davis, and Hanley (2007). When naming pictures of familiar objects, adults are slowed by the presence of the orthographic form of a word that is not the name of the object, e.g., naming a picture of a cat in the presence of the word MOP. This demonstrates the automaticity of orthographic processing. Naming is further slowed if the word is semantically related to the object, e.g., naming a cat in the presence of the word SHEEP, indicating that semantic information is also processed extremely rapidly. This effect is known as picture-word interference. Clay et al. taught adults novel names for novel objects from three categories: fruit, vehicles, and items of clothing. The novel objects were both pictured and described, e.g., KOSLA = a bitter and spiky fruit. Participants then completed a picture-word interference task. Clay et al. found that the presence of a trained novel word interfered with the recognition of a different trained object, whereas the presence of an untrained novel word did not. This interference effect was present immediately after training and did not increase when participants were re-tested one week later. This was taken as evidence that the

orthographic forms of trained novel words were processed automatically immediately after training. Interestingly, immediately after training, the presence of a trained word from the same category did not cause greater interference than an unrelated trained word, but this semantic interference effect did emerge when participants were re-tested one week later. This suggests that the semantic properties of newly learned orthographic forms may require consolidation to be processed automatically.

Overall these novel word learning experiments suggest that consolidation, and possibly overnight sleep, may be important in phonological, orthographic, and semantic learning. They also provide some evidence to suggest that the integration of these different types of information into existing knowledge may follow different time courses. However, the use of different outcome measures in the studies described above makes it difficult to draw firm conclusions on this point. The experiments reported in this thesis did not directly examine the role of sleep in novel word learning. In Experiment 3, old-new decision latencies were significantly faster following a second day's training, but this confounds additional practice with overnight sleep. Future work should directly examine the effect of sleep vs. extended training on learning a novel orthography. As it is inherently more difficult to learn new symbols as well as sounds, it may be that the integration of novel orthographic forms into existing knowledge will follow a more protracted time course than that observed in the above described experiments.

Somewhat contradictory evidence concerning lexicalisation of new words is provided by research on first (L1) vs. second language (L2) processing. Gollan, Forster,

and Frost (1997) found that in Hebrew-English bilinguals, masked L1 translation primes facilitated lexical decisions to L2 targets, but that this did not hold for L2 primes and L1 targets. This was replicated by Jiang (1999) in Chinese-English bilinguals. Jiang and Forster (2001) suggested that this asymmetry might indicate that L2 words are represented in episodic memory not lexical memory. This would account for why L2 primes did not influence performance in a lexical task. This was supported by a series of experiments demonstrating that L2 primes facilitated responding to L1 targets in an episodic task, but not a lexical task. In the episodic task, Chinese-English bilinguals studied a list of 32 Chinese (L1) words and then had to indicate in a test phase whether presented words had been studied. In this test phase, masked English (L2) translation primes decreased RTs to studied words. In contrast, and replicating previous research, L2-L1 priming did not occur in a lexical decision task. These findings demonstrate that L2 words are able to influence episodic but not lexical processing, suggesting that L2 words are not integrated into lexical memory.

This conclusion seems at odds with the findings of Gaskell and Dumay (Dumay & Gaskell, 2007; Gaskell & Dumay, 2003) and Bowers and colleagues (Bowers et al., 2005; Clay et al., 2007) which suggested that new words are incorporated into the existing lexicon on the same day, or following overnight sleep. However, one important difference is that in these novel word learning studies, stimuli were presented as new members of the existing language, whereas this tends not to be the case when acquiring a second language. This difference might influence the integration of newly learned items.

It thus seems important to consider whether, in the experiments reported in this thesis, newly learned words were treated as new members of the existing language or as a different language. In Experiment 1, participants learned new words written in English orthography and items were assigned novel meanings rather than being presented as translations of existing words. It therefore seems likely that participants treated these words as belonging to their existing language. This may have contributed to the fact that the novel words in Experiment 1 were processed similarly to real words. In contrast, in Experiments 2 to 5, participants were instructed that they would be learning a new language, and the fact that items were written in a novel script will have reinforced this. Thus, it may be that newly learned words in the artificial orthography should function more like words in a second language. On this view, perhaps we should not expect them to be integrated into existing lexical memory? However, as in Experiment 1, trained items in Experiments 4 and 5 were assigned novel meanings rather than being presented as translations of existing words. Thus, there were both similarities to and differences from second language learning in these artificial orthography experiments. Overall, it seems that there is much to be learned about how learning and subsequent processing of newly learned items resembles that of existing items, in both novel word and artificial orthography learning paradigms.

If it is possible that newly learned items in the artificial orthography were represented differently from existing words, do the results of this thesis inform us about learning to read in natural alphabetic languages? As described earlier, lexical-like effects of frequency and consistency emerged from exposure to the orthography, influencing

learning, discrimination, and generalization. Furthermore, the results of Experiment 5 replicated those of a novel word learning study which used English orthography to investigate the influence of semantic vs. phonological familiarity on orthographic processing (McKay et al., 2008). Thus, in the artificial orthography paradigm, effects of frequency, consistency, and semantics on orthographic processing mirrored those seen in English orthography. Without wishing to claim that this method mimics all of the processes involved in lexical learning, it does provide a high degree of experimental control and generates data that reliably simulate typical reading processes.

Operationalizing frequency and consistency

The artificial orthography paradigm was developed in response to several problems identified with natural language research concerning the measurement and control of lexical variables such as frequency, consistency, age-of-acquisition, and semantics. This section considers whether the frequency and consistency manipulations, used in Experiments 2 to 5, were appropriate for considering the questions of interest. The same question is asked of the semantic manipulation, used in Experiments 1, 4, and 5, in a later section.

Frequency. Research on natural languages has focused on token frequency manipulations. For example, high frequency words are read faster than low frequency words, particularly when they contain atypical spelling-sound mappings (Jared, 2002; Paap & Noel, 1991; Seidenberg et al., 1984). The influence of token frequency is also reflected in the Dual Route Cascaded (Coltheart et al., 2001) and triangle (Harm & Seidenberg, 2004; Plaut et al., 1996) models of reading aloud. In contrast, the artificial

orthography incorporated a type frequency manipulation: in each stimulus set there were three different vowel types (consistent, inconsistent-conditioned, inconsistent-unconditioned), and each had a high frequency version, which occurred in eight words, and a low frequency version, which occurred in four words.

Kessler, Treiman, and Mullenix (2003) suggested that "... The more often we see something, the better we should learn it and the more readily we should recall it..." (p. 11). This makes it clear that token frequency should influence learning. However, they also comment that "... On the other hand, we clearly categorize information. The fact that we see the word *of* extremely often does not necessarily drive adults to consider /v/ as a serious contender for the pronunciation of *f* in other words..." (p. 11). This illustrates why type frequency (the fact that *f* is pronounced /f/, not /v/, in most words) is also important. Kessler et al. then went on to examine how well type and token frequency measures of orthographic consistency predicted adult naming behaviour. Results indicated that token frequency accounted for additional variance in naming speed over and above type frequency but that "... the differences between approaches using type counts and the frequency-weighted counts were small..." (p. 25). Thus, in the artificial orthography, it was predicted that a type manipulation would be sufficient to examine the influence of frequency on learning and subsequent processing.

The results of Experiments 2 to 5 supported this prediction. Across experiments, reading accuracy was higher for trained items containing high frequency vowels than for those containing low frequency vowels, particularly when items had inconsistent character-sound mappings. This mirrors token frequency effects reported in natural

language studies (e.g., Jared, 2002; Weekes et al., 2006). Furthermore, type frequency effects were also present in generalization: novel items were read more accurately when they contained vowels that were highly frequent during training. It is clear why this should be the case: greater experience with a vowel pronunciation should increase reading accuracy for items containing that vowel. Thus, a type frequency manipulation was sufficient to examine learners' sensitivity to the frequency characteristics of the language and generated reading accuracy data which closely resembled findings from natural languages.

In Experiments 3 and 4, items containing high frequency vowels were recognised more quickly in old-new decision, simulating effects seen in typical lexical decision (Balota et al., 2004; Cortese & Khanna, 2007). However, in Experiments 3 and 4, the facilitative effect of type frequency was not reflected in old-new decision accuracy. In contrast, in Experiment 5 frequency had a pervasive effect on old-new decision accuracy and a less clear influence on RTs. Thus, frequency effects in old-new decision were somewhat unreliable across experiments. It is possible that the use of a type rather than token frequency manipulation contributed to this. High frequency words are recognised more quickly in lexical decision because they have been experienced often. In the high frequency items in the artificial orthography however, only the vowels and not the whole items had been experienced more often. Type frequency effects in old-new decision might therefore be expected to be weaker than token frequency effects in typical lexical decision. Overall, the type frequency manipulation used in the artificial orthography was sufficient for the purpose of examining sensitivity to the statistics of

the language during learning. However, the interplay between type and token frequency and their influences on different outcome measures represent important areas for future research.

Consistency. The inadequacy of the consistency manipulation used in Experiment 1 was discussed in detail in Chapters 2 and 3. This discussion will therefore focus on the manipulation of consistency in those experiments using the artificial orthography paradigm (Experiments 2 to 5). As discussed earlier, this manipulation provided the first conclusive demonstration that knowledge of context-sensitive character-sound mappings can be extracted through exposure to the orthographic form of whole words and their pronunciations. Furthermore, it was also shown that the predictability of character-sound mappings influenced reading aloud at the end of training and generalization in a way that mirrored observations from English orthography (Jared, 2002; Kessler & Treiman, 2001; Weekes et al., 2006). However, there are several aspects of the consistency manipulation that warrant further consideration. The first concerns the saliency of conditional pronunciations of inconsistent vowels. To illustrate: of the 36 items in stimulus set 1, eight began with /vi/ (high frequency inconsistent-conditioned) and four began with /zəʊ/ (low frequency inconsistent-conditioned). These frequently repeated onset-vowel combinations may well have made inconsistent-conditioned items highly salient. Although participants were not explicitly aware of the rules governing inconsistent vowel pronunciations, this high salience may explain why items containing high frequency inconsistent-conditioned vowels were learned as easily as items containing high frequency consistent vowels. It is clear that there are differences

between this restricted item set and relatively salient consistency manipulation, and the ever increasing and far more variable set of words that learners of natural alphabetic languages must cope with. Future work should use larger training sets and decrease the salience of particular item types. It is likely that such an adaptation would increase the dominance of consistent over inconsistent items and that longer training would be necessary for learners to develop context-sensitivity.

A second issue is that vowel pronunciations were entirely deterministic: if a learner extracted the conditional rules they could achieve 100% accuracy in training and generalization. Although the vast majority of learners did not develop explicit knowledge of this kind (as evidenced by less than perfect generalization), the deterministic nature of vowel pronunciations in the artificial orthography was less complex than the probabilistic relationships between letters and sounds in English orthography. Artificial language learning experiments have demonstrated that learners are sensitive to probabilistic relationships and that their subsequent language use reflects the statistics of their learning environment (e.g., Wonnacott et al., 2008). Thus, it seems likely that learners of an artificial orthography would develop and use probabilistic spelling-sound knowledge. The incorporation of non-deterministic conditional relationships between characters and sounds represents an exciting direction for subsequent artificial orthography learning research.

Another possible adaptation to the consistency manipulation would be to condition vowel pronunciations according to codas as opposed to onsets. Native English speaking participants would probably find such relationships easier to learn because in

English orthography, vowel pronunciations are affected more often by final than initial consonants (Treiman et al., 1995). However, it might also be the case that prior knowledge could interfere with learning, particularly if coda conditioning rules were not of the same nature as those that exist in English. Overall, manipulating the predictability of character-sound mappings represents one of the most exciting potentials for artificial orthography paradigms.

A further consideration is that although consistency was entirely controlled within the artificial orthography, it is possible that the feedback consistency of the sounds in English impacted on learning. For example, the body /əʊt/ can be spelled in multiple ways in English, e.g., OAT, OTE, whereas /ʌt/ can only be spelled in one way, UT. This might have influenced the ease with which items containing such bodies were learned. It is not entirely clear which of these cases would prove most difficult. It might be that it is easier to learn new associations for bodies that have multiple sound-spelling mappings in English because they are already flexible. However, it might also be that the existence of multiple mappings in English increases confusion. It would be extremely difficult to eliminate such feedback consistency effects in this paradigm. The only way to remove this confound would be to teach novel sounds as well as novel characters. However, this would pose considerable additional challenges which would be likely to interfere with learning in ways that might prove difficult to control/assess.

The role of phonological and semantic familiarity in learning to read aloud

Experiment 1 provided some evidence to support the idea that semantics plays a role in reading aloud. It also suggested that familiarity with the phonological forms of items may be important. However, experiments such as this, in which participants read novel words written in a familiar alphabetic script, provide poor control over lexical variables such as spelling-sound consistency and orthographic age-of-acquisition. They also fail to address the role of semantics in the early stages of learning to read. In Experiments 4 and 5, the artificial orthography paradigm was therefore used to examine the impact of semantics on orthographic learning. Experiments 2 and 3 demonstrated that this method provides a high degree of experimental control and produces strong and reliable effects of frequency and consistency on learning. This made it an ideal tool for investigating semantic effects on word reading processes. A particular problem for natural language research is the difficulty of separating potential influences of phonological and semantic familiarity. It was possible to dissect the effects of these variables in Experiments 4 and 5 by manipulating whether participants were familiarised with item sounds (lexical phonology) or sounds plus meanings (semantics), prior to learning the orthography.

However, a further issue arose concerning whether semantic vs. lexical phonology pre-exposure should be manipulated between or within subjects. A previous novel word learning study by McKay et al. (2008, Experiment 2) used a within subject semantic manipulation in which participants learned definitions for half the items (semantic pre-exposure) and only the sounds of the other half of the items (lexical

phonology pre-exposure). This design meant that participants might have focused on the semantic learning task and paid minimal attention to the lexical phonology items. To avoid this problem, Experiment 4 manipulated semantic vs. phonological familiarity between subjects, by pre-exposing 16 participants to novel word sounds and 16 to novel word sounds plus a novel object referent, intended to represent semantics. Both pre-exposure conditions provided equivalent benefit for reading accuracy during training, relative to a no pre-exposure condition provided by Day 1 of Experiment 3. Importantly, by the end of training, this benefit was specific to items containing low frequency inconsistent character-sound mappings. Lexical phonology and semantic pre-exposure also increased the speed with which trained items could be discriminated from untrained items. In this experiment then, phonological familiarity produced the same facilitation that natural language research (and McKay et al. (2008)) has attributed to semantic variables. The results of Experiment 4 therefore contrasted with those reported by McKay et al. and suggested that when attention is equated during pre-exposure, semantic familiarity does not support orthographic learning over and above the beneficial effects of phonological familiarity.

An obvious caveat to this argument is that it is likely that novel objects induced semantic representations that were weaker than those induced by the novel definitions used by McKay et al. (2008), and indeed than those of words in natural languages. In Experiment 5, McKay et al.'s novel definitions were therefore used to instantiate semantics at pre-exposure. This was again compared to lexical phonology pre-exposure, but using a within subject design, as in McKay et al.'s experiment. The results of this

experiment could then be directly compared to those reported by McKay et al. Importantly, however, it was also possible to make comparisons between the within subject semantic and lexical phonology conditions in Experiment 5, and the between subjects no pre-exposure and lexical phonology conditions from Experiments 3 and 4. This enabled direct assessment of the benefit of semantics, instantiated with definitions, in both a within and a between subjects design.

A somewhat complex picture emerged. Within Experiment 5, semantic pre-exposure provided greater benefit than lexical phonology pre-exposure during orthography training. At the end of training, this benefit was specific to items containing low frequency inconsistent vowels. Post-training discrimination of trained from untrained items was also more accurate following semantic pre-exposure. These findings mirror those of McKay et al. (2008). In addition, semantic pre-exposure conferred greater benefit during training than the lexical phonology condition from Experiment 4. Thus far, the results support the idea that semantic familiarity provides greater benefit for orthographic learning than phonological familiarity when instantiated with definitions rather than novel objects.

However, further analyses pointed to an alternative conclusion. By the end of training, reading accuracy was equivalent in the lexical phonology condition from Experiment 4 and the semantic condition from Experiment 5. Both pre-exposure conditions improved end of training reading accuracy for items containing low frequency inconsistent-unconditioned vowels, relative to no pre-exposure. Discrimination was also equivalent in these two conditions and in fact, across

experiments there was no evidence that either phonological or semantic familiarity enhanced discrimination. Furthermore, end of training and old-new decision accuracy were lower following lexical phonology pre-exposure in Experiment 5 than in the equivalent (between subjects) condition in Experiment 4. This means that the advantage of definitions over lexical phonology pre-exposure observed in McKay et al.'s (2008) experiment may to some extent have been an artefact of their within subject semantic manipulation, rather than arising from the semantic content of the definitions.

It is worth noting that some of the above described effects, in particular those that constituted three-way interactions, were significant across items but not across subjects. For example, a three-way interaction was reported between pre-exposure, frequency, and consistency in the end of training analysis which compared the semantic condition from Experiment 5, the lexical phonology condition from Experiment 4, and the no pre-exposure condition provided by Day 1 of Experiment 3. This revealed that semantic and lexical phonology pre-exposure provided equivalent benefit for items containing low frequency inconsistent-unconditioned vowels, relative to no pre-exposure. This interaction was only significant in the items and not the subjects analysis. However, the pattern of pre-exposure effects was consistent across Experiments 4 and 5. For example, in both experiments, pre-exposure increased accuracy across items early in training with effects becoming restricted to inconsistent low frequency items by the end of training. Thus, it seems likely that theoretically important interactions would reach significance in both subjects and items analyses with the inclusion of more participants.

To summarise, comparisons across experiments demonstrated that familiarity with word meanings benefited learning to read aloud, over and above the benefit conferred by familiarity with word sounds. However, specific semantic effects on end of training performance on items containing low frequency inconsistent vowels were not observed. Instead, lexical phonology and semantic pre-exposure conferred the same benefit, relative to a no pre-exposure baseline. Previous natural language research that has not considered the role of phonological familiarity may therefore have overestimated the role of semantics in word reading.

Tentative support for this position is provided by a recent experiment by Nation and Cocksey (2009). They found that although 8-year-old children were better able to read words that were phonologically familiar, no additional benefit resulted from being able to define a word. Importantly, the predictive relationship between phonological familiarity and reading aloud only held for irregular words, not regular words. This is consistent with the findings of Experiments 4 and 5 in which lexical phonology pre-exposure provided equivalent benefit to semantic pre-exposure for end of training performance on items containing low frequency inconsistent vowels.

The problem of instantiating semantics

The finding that phonological and semantic familiarity conferred equivalent benefit to end of training performance contrasts with previous natural language research suggesting that imageability (a semantic variable) influences adults' naming accuracy speed for low frequency inconsistent words (Frost et al., 2005; Strain et al., 1995). In this thesis it has been argued that such demonstrations are open to multiple

interpretations due to the natural correlations between psycholinguistic variables, in particular, between imageability and age-of-acquisition (AoA) (Cortese & Khanna, 2007; Ellis & Monaghan, 2002; Monaghan & Ellis, 2002). However, Experiment 5 did not directly manipulate imageability and although participants learned the definitions well, it is likely that semantic representations were not as strong or rich as those of high imageability familiar words. Potentially, therefore, they may have been too weak to drive specific semantic influences on orthographic processing at the end of training.

Given the difficulties of investigating the influences of semantic variables in natural languages, future research could use the artificial orthography paradigm to precisely control (or vary) both phonological and orthographic AoA, and manipulate the strength of semantic pre-exposure, perhaps by using novel definitions that vary in imageability. This was attempted in Experiment 1, in which novel words were written in English orthography. However, in this experiment, high imageability definitions did not confer greater benefit than low imageability definitions. It was suggested that, in this experiment, the advantage of semantic over lexical phonology pre-exposure (seen only in reading accuracy during training) may have been a result of increased attention rather than the semantic content of the definitions. Again it is likely that extended training will be necessary for variations in semantic stimuli to influence word reading in an artificial orthography paradigm.

The difficulty of operationalizing semantics poses significant problems for future word learning research. In an experimental paradigm it is very difficult to induce semantic representations of novel words that are as strong or rich as those of words in

natural languages. Words in natural languages vary along multiple dimensions (e.g., meaningfulness, number of associated words, number of semantic features) all of which have been found to influence orthographic processing, over and above other variables such as frequency and consistency (Balota et al., 2004; Pexman et al., 2008). In addition, Bolger, Balass, Landen, and Perfetti (2008) have shown that adults learn the meanings of new words more successfully when they are presented in multiple contexts, as opposed to repeated presentations of the same context. Furthermore, the context in which we experience new words impacts on orthographic processing. Adelman, Brown, and Quesada (2006) demonstrated that measures of contextual diversity (the number of different contexts in which a word has been experienced) accounted for additional variance in adults' naming and lexical decision latencies after word frequency was taken into account. This contextual variation is not provided by experimental studies in which novel word meanings are instantiated using a single definition. Thus, such studies neglect an important characteristic of our experience with words in natural languages. It is clear that future word learning studies should aim to provide richer experiences with novel word meanings.

The effects of phonological/semantic familiarity at different stages of learning

Effect of pre-exposure on orthography training. As described above, comparisons across Experiments 3 to 5 demonstrated that familiarity with word meanings influenced the early stages of learning to read over and above familiarity with word sounds. However, specific semantic facilitation when reading items containing low frequency inconsistent vowels at the end of training was not observed. At the end of Chapter 6 it

was suggested that higher proficiency on the artificial orthography might be required for these specific semantic effects to emerge. This suggestion was based on analyses of the entire orthography training phase which indicated that the advantage of semantic over phonological familiarity was greater in the middle than the early stages of training. Perhaps then with extended training, specific semantic effects on end of training performance on items containing low frequency inconsistent vowels would be evident. Earlier it was suggested that a further effect of extended training might be to increase the extent to which trained items were processed similarly to existing words. It seems possible that these two factors might be related. If newly learned words start to function more like real words, the effect of semantics on orthographic processing might strengthen.

The artificial orthography paradigm enabled the effects of phonological/semantic familiarity to be assessed in the early stages of learning to read. This generated the particularly interesting result that pre-exposure improved reading accuracy across items at the beginning of training, with facilitation becoming restricted to items containing low frequency inconsistent vowels by the end of training. This result was obtained in both Experiments 4 and 5 and suggests that facilitative effects of phonological/semantic familiarity on orthographic processing may change over time. This idea has not received much attention in natural language research because most studies have examined semantic effects at fairly high levels of reading skill.

As discussed in Chapter 5, some notable exceptions provide support for the idea that semantic/phonological familiarity may influence both consistent and inconsistent

word reading in the early stages of learning. Laing and Hulme (1999) found that beginning readers were better able to learn printed abbreviations for spoken word targets (e.g. bgn for bacon) when targets were highly imageable irrespective of the phonological similarity of the cue to the target. Similarly, Nilsen and Bourassa (2008) found that children learned to read regular words more easily than irregular words and concrete words more easily than abstract words, but again these variables did not interact. Furthermore, Ouelette and Fraser (in press) found that providing word meanings benefited orthographic learning for regular/consistent words in a sample of somewhat older children age 9.5 years. Together with the results of Experiments 4 and 5, these experiments suggest that in the early stages of learning to read, semantic/phonological familiarity may support the learning of orthography-phonology mappings for all words, not just those with atypical mappings between spelling and sound.

A point to note is that none of these studies explicitly controlled for attention to phonological form or directly examined the role of phonological vs. semantic familiarity. For example, in Nilsen and Bourassa's (2008) experiment, children may have found concrete words more interesting and paid more attention to how they sounded and were written. Thus, although these experiments suggest that semantic familiarity supports the early stages of learning to read, irrespective of spelling-sound typicality, they leave open the possibility that semantic effects were driven by enhanced attention rather than semantic properties of the stimuli. This is supported by an experiment by McKague et al. (2001) in which 6- to 7-year-old children learned to read novel words

with consistent spelling-sound mappings. Orthographic learning benefited from prior phonological familiarity, but semantic familiarity provided no further benefit. This supports the idea that phonological rather than semantic familiarity supports the early stages of learning to read words, and again suggests that this effect operates irrespective of spelling-sound consistency.

Effects of pre-exposure on generalization. An unexpected effect of pre-exposure, observed in both Experiments 4 and 5, was that it reduced generalization performance. Furthermore, this reduction in generalization was most pronounced for items containing low frequency inconsistent vowels, exactly those that were boosted by pre-exposure during training. It was suggested that because pre-exposure resulted in stronger phonological representations of training items, this might have enabled participants to read training items by recalling that they were a valid word in the language. While this improved reading accuracy for specific items during training, it may have reduced the necessity to extract information about the influence of consonants on vowel pronunciation. In turn, this would have been detrimental for generalization. Research on orthographic learning in children supports this idea, suggesting that while contextual support can improve reading accuracy, it may not be beneficial for the retention of spelling-sound mappings (Landi et al., 2006). This research also found that learning in context had a less negative effect on retention for more advanced readers. This is encouraging as it suggests that with extended orthography training, generalization would not be reduced following pre-exposure.

Relationship to existing models of reading aloud

Three points seem particularly important to consider with regards to how the results presented in this thesis fit with existing models of reading aloud: the extraction of character-sound mappings, the representation of item-specific phonological and semantic information, and the time course of phonological/semantic effects on learning.

Character-sound mappings. The DRC model of reading aloud (Coltheart et al., 2001) suggests that a non-lexical route stores rules for converting graphemes (letters or letter sequences) into phonemes and a lexical route stores whole-word orthographic forms and their pronunciations. As described in Chapter 1, the grapheme-phoneme conversion rules (GPCs) that characterise the non-lexical route convert a grapheme to its most common single phoneme pronunciation. Thus, GPCs are largely insensitive to context, i.e., what the other letters in a word are. For example, the DRC does not take account of the fact that, although the most frequent pronunciation of OO is /u/, OOK is most commonly pronounced /ʊk/. Thus, words such as BOOK, TOOK, and LOOK would have to be stored as whole-word forms and read by the lexical route, despite the fact that OOK has a predictable pronunciation.

It is unclear how GPCs are learned because Coltheart et al. (2001) chose not to examine learning on the grounds that "... unless the learning procedure itself is known to be psychologically real, it may not be able to learn what people learn..." (p. 216). However, the DRC does predict that where ambiguities exist, learners should use the most frequent pronunciation of a character when reading novel words. The results of Experiment 2, and replications provided by Experiments 3 to 5, do not support this

prediction. Instead, when reading generalization items containing inconsistent vowels, learners used both high and low frequency pronunciations. Furthermore, pronunciations used in generalization were to a large extent context-sensitive, i.e., learners used the conditioned and unconditioned pronunciations following appropriate consonant onsets.

Thus, generalization behaviour seems to be inconsistent with the way in which spelling-sound knowledge is represented in the DRC. Instead it aligns far better with the representation of orthography-phonology mappings in the triangle model of reading aloud (Harm & Seidenberg, 2004; Plaut et al., 1996). This model has no inbuilt GPCs and learns to read by being presented with the orthographic form of a word, producing a pronunciation attempt and receiving the correct pronunciation as feedback. Importantly, the model develops implicit knowledge of context-sensitive spelling-sound mappings, pronouncing a nonword such as VOOK to rhyme with BOOK, TOOK, and LOOK, rather than SPOOK, and a nonword such as VOOT to rhyme with BOOT, ROOT, and LOOT, rather than FOOT. This seems far more similar to the context-sensitive character-sound mappings that developed through exposure to the artificial orthography and which were used successfully in generalization. Furthermore, post-experiment debriefing indicated that only three participants (from a total of 92 across experiments) were explicitly aware of any of the rules governing inconsistent vowel pronunciations. Thus, the knowledge they developed of character-sound mappings appears to be implicit, again in line with the triangle model.

The consistency manipulation incorporated in the artificial orthography was not specifically designed to pit learning of GPCs against context-sensitive spelling-sound

mappings. This represents a potential direction for future research. There is evidence to suggest that children use a combination of GPCs and sensitivity to consistency in reading (Stuart et al., 1999; Treiman et al., 1990) and spelling (Treiman & Kessler, 2006). However, it is difficult to assess the extent to which the contribution of GPCs (and indeed consistency) reflects the influence of direct teaching. Therefore, a further informative extension would be to include explicit character-sound mapping instruction in training. Powell, Plaut, and Funnell (2006) compared learning in a version of the triangle model following training with and without GPCs and found that GPC instruction did improve performance. However, this work did not directly contrast learning of different word types, i.e., regular vs. irregular, consistent vs. inconsistent, thus it is not possible to evaluate why GPC training was beneficial. The control over exposure to lexical statistics provided by artificial orthography paradigms makes them an ideal tool for separating the predictions made by the DRC and triangle models of reading aloud, and also for investigating the impact of explicit vs. implicit learning of spelling-sound mappings.

Phonological or semantic familiarity? The triangle model suggests that semantic representations influence the computation of phonology from orthography for known words. However, the results of Experiments 4 and 5, and indeed those of Experiment 1, suggested that phonological (as well as semantic) familiarity plays a role in learning to read aloud. If this is the case, the triangle model would need to be modified such the production of phonology from orthography is supported not only by semantics, but by a mechanism that binds together the phonemes in known words in a qualitatively

different way from that which occurs for common sub-lexical combinations of phonemes. At present, the triangle model does not represent lexical, or whole-word, phonology in any way.

This might seem more akin to the lexical route of the DRC which stores whole-word orthography-phonology representations for known words. However, there was little evidence that pre-exposure to lexical phonology improved participants' ability to read inconsistent-unconditioned items by inducing word-level correspondences between orthographic and phonological forms. Participants had great difficulty with old-new decisions suggesting that strong word-level orthographic representations did not develop. Furthermore, all vowel types, including inconsistent-unconditioned, were used in generalization. This again indicates that items containing such vowels were not learned in a whole-word manner. Thus, a modification to the triangle model seems more appropriate.

The phonological attractor network incorporated in a later version of the triangle model increased within-word phoneme binding by improving the model's "... knowledge of the segmental structure and constraints on sequences of phonemes..." (Harm & Seidenberg, 1999, p. 43). However, these recurrent connections did not necessarily induce the whole-word phonological representations which the results of Experiments 4 and 5 suggest to be important. In fact, a reconsideration of Plaut et al.'s (1996) implementation of the triangle model suggests an alternative way in which whole-word phonological input could be provided. As described in Chapter 1, in this implementation of the triangle model, so-called 'semantic support' operated by providing "... additional

input to the phonological units, pushing them towards their correct activations...” (p. 95). This input could quite easily be re-cast as arising from the possession of item-specific phonological representations.

Later versions of the triangle model have used distributed, similarity based representations to encode semantics (Harm & Seidenberg, 2004). Unfortunately though there has been no discussion of how these semantic representations impacted on word reading, relative to the impoverished representations used by Plaut et al. (1996). Instead, discussion focused on reading comprehension. Future modelling work should incorporate both Plaut et al.’s original “... additional input to the phonological units...” or perhaps a modified version of Harm and Seidenberg’s (1999) phonological attractor network (to capture effects of item specific phonological familiarity), and richer semantic representations. Such work could then examine the relative benefit of these two sources of information over the course of training.

The development of effects of phonological/semantic familiarity. Experiments 4 and 5 demonstrated that early in training, pre-exposure improved reading accuracy across training items. Pre-exposure effects only became restricted to items containing low frequency inconsistent-unconditioned vowels at the end of training. These findings clearly demonstrate that the influence of phonological/semantic familiarity on orthographic processing, and indeed the distinction between consistent and inconsistent words, is not all-or-nothing. This is compatible with the triangle model framework, in which all words rely on the same sets of representations and the “...

division of labor..." (Plaut et al., 1996, p. 58) between such representations (phonological, orthographic, semantic) is graded and can change over time.

Coltheart et al. (2001) proposed that the lexical route of the DRC could be further subdivided into a semantic and a non-semantic pathway. This would allow word meanings to play a role in word pronunciation. However, as the semantic lexical route has not been implemented, it is difficult to make explicit predictions as to how a semantic contribution to reading aloud might operate. Furthermore, the DRC is explicitly non-developmental and thus cannot account for changes in the influence of semantic/phonological familiarity over time. Finally, in the DRC, unfamiliar words are read by the non-lexical route which operates solely on the basis of GPCs, i.e., without support from item-specific phonological/semantic representations. It is therefore difficult to see how this model could account for the pervasive effects of pre-exposure to phonology/semantics at the very beginning of training, as observed in Experiments 4 and 5.

However, the current findings also challenge Plaut et al.'s (1996) assumption that semantic knowledge only makes a substantial contribution to reading aloud once orthography-phonology mappings have developed to some degree. Plaut et al. offered two reasons why this should be the case. First, because early reading instruction focuses heavily on learning letter-sound rules, and second, because letter-sound mappings are structured whereas the links between forms and meaning are largely arbitrary. On this basis, Plaut et al. increased the input from semantics over the course of training and there was no opportunity for semantic influences in the earliest stages of learning. The

results of Experiments 4 and 5 suggest that Plaut et al.'s assumption is not valid. Instead, our system seems to capitalize on the fact that links between phonology and semantics are in place before we learn to read and uses this information in the early stages of orthographic learning.

Further ways to assess orthographic learning

In the experiments reported in this thesis, reading accuracy was assessed but speed was not. At the end of Chapter 6 it was suggested that it might be possible to measure naming speed if participants received extended training with the artificial orthography. Extended training would be necessary because reading aloud was relatively laborious following the 30-45 minutes orthography training received in Experiments 2 to 5. Potentially, naming speed might be more sensitive to semantic variables than naming accuracy. In line with this, Strain et al. (2002) found imageability to be a significant predictor of naming latency but not errors, although as discussed in Chapter 2, Strain et al.'s data are compromised in various ways. However, in Trudeau's (2006) novel word learning study, the imageability of novel definitions again influenced subsequent naming speed but not accuracy.

This thesis focused on reading aloud but it would also be possible to examine spelling in the artificial orthography. Although many of the same factors influence reading and spelling, research suggests that they may not be influenced by exactly the same factors at exactly the same times (Bradley & Bryant, 1979). Caravolos, Kessler, Hulme, and Snowling (2005) investigated consistency effects in young children's spelling. They found that spelling-sound regularity and word frequency predicted the number of

children who could spell a word correctly but that consistency did not. This stands in contrast to the work of Treiman et al. (2006) which found that 6-year-old children are sensitive to consistency when reading aloud novel words. Thus, it is possible that learners might differ in their ability to demonstrate sensitivity to orthographic consistency in reading and spelling tasks. The addition of spelling as an outcome measure would therefore add to our understanding of the development of spelling-sound knowledge.

To further investigate learning to spell it would be of interest to introduce a feedback consistency manipulation to the artificial orthography, i.e., to have sounds represented by more than one character. Treiman and Kessler (2006) found that children are sensitive to consonant context when spelling vowels. Of particular interest is their finding that children used onsets to predict vowel spellings at the age of 8 to 9 years, but did not reliably use codas to predict vowel spellings until age 11. Again this indicates differences between the influence of context-sensitive spelling-sound mappings on reading and spelling because children as young as 6-years-old are influenced by consonantal context when pronouncing vowels (Treiman et al., 2006). Furthermore, Treiman et al. (2006) demonstrated that in reading aloud, sensitivity to onset and coda conditioning emerges at around the same time. Thus, incorporating a feedback consistency manipulation and investigating spelling in the artificial orthography would provide a further way to assess sensitivity to context in orthographic learning.

Research has demonstrated that feedback consistency also influences *auditory* lexical decision. Data from English, (Chereau, Gaskell, & Dumay, 2007), French (Ziegler & Ferrand, 1998), and Portuguese (Ventura, Morais, Pattamadilok, & Kolinsky, 2004) adults has found that the discrimination of words from nonwords in the auditory modality is slower when words contain a body that could be written in more than one way, e.g., in English, BIRD could be spelled, BURD, or BERD, than when they contain bodies that can have only one spelling, e.g., FISH. This demonstrates that the mappings between orthography and phonology are strong and reciprocal such that orthographic inconsistencies influence word recognition, even in a task in which orthography is not present. If a feedback consistency manipulation was incorporated into the artificial orthography and it influenced auditory old-new decision, this would provide some evidence to suggest that items in the artificial orthography are processed in a similar way to words in natural languages.

The current experiments used behavioural measures to assess orthographic learning. The arguments made on the basis of these experiments might be strengthened by incorporating neuroimaging methods. For example, Experiments 3 and 4 demonstrated that participants developed sensitivity to spelling-sound consistency and frequency through exposure to whole-word forms. The same influences of frequency and consistency were then observed in both generalization and recognition of whole items. These findings support the idea that item-specific and general spelling-sound knowledge is represented in a single system, as suggested by the triangle model of reading aloud (Harm & Seidenberg, 2004; Plaut et al., 1996). In contrast, previous

neuroimaging research has demonstrated that word (item-specific) and nonword (generalization) brain activation differs (Jobard, Crivello, & Tzourio-Mazoyer, 2003). However, this may not be because words require item-specific knowledge and nonwords require sublexical spelling-sound knowledge, but instead may reflect the fact that words recruit additional semantic processing regions (Binder, Medler, Desai, Conant, & Liebenthal, 2005).

If neuroimaging techniques were combined with an artificial orthography paradigm it might be possible to separate these two suggestions. This is because trained (word) vs. untrained (nonword) item processing could be examined in a situation in which trained items had no meaning. If item-specific and generalization behaviour depend on separable representations, different brain regions should be active during recognition of trained items and reading of untrained items. In contrast, the same regions should be active if these processes depend on the same underlying representations. Furthermore, if a semantic vs. lexical phonology pre-exposure phase was introduced, it would be possible to assess whether brain regions that are typically active during semantic processing, e.g., the angular and medial temporal gyri, were recruited during subsequent orthographic learning. This would provide a new way to investigate the suggestion made in Experiments 4 and 5, that semantic and phonological familiarity influence all words in the early stages of learning to read, with effects becoming restricted to inconsistent words with increasing proficiency.

Final comments

Natural language research and novel word learning experiments using English orthography suffer from methodological problems due to difficulties in measuring and controlling for the many variables that influence orthographic processing. Experiment 1 provided a clear demonstration of these problems and, in response, an artificial orthography paradigm was developed in Experiment 2. This paradigm offered one solution to these issues, providing a greater degree of experimental control than previous experiments, and enabling direct examination of the factors influencing orthographic learning. Using this artificial orthography paradigm, Experiments 2 and 3 provided the first clear demonstration that learners can extract context-sensitive character-sound mappings from exposure to whole-word orthographic forms (and corresponding pronunciations) and use these productively. Furthermore, effects of frequency and consistency on learning and generalization emerged which mirrored those seen in English orthography. This showed that learners are sensitive to lexical properties that are implicit in the language environment and validated the paradigm for investigating further theoretical questions. This was achieved in Experiments 4 and 5 which examined the role of semantics in orthographic learning. Results suggested that both semantic and phonological familiarity benefit learning to read.

Overall, findings were broadly compatible with the "... division of labor..." embodied in the triangle model of reading aloud (Plaut et al., 1996, p. 58). However, a key area for the future will be to examine how extended orthography training and the provision of richer semantic information influence the relative benefits that

phonological and semantic familiarity provide for orthographic learning. Of equal importance will be to investigate how newly learned orthographic forms are represented in memory and how this changes over time and with increasing experience. Overall, the experiments reported in this thesis have shown that an artificial orthography paradigm can be used to investigate orthographic learning and generalization. In future, artificial orthography experiments should be complemented by computational modelling and neuroimaging research, and by longitudinal and training experiments in developing readers.

REFERENCES

- Adelman, J. S., Brown, G. D. A., & Quesada, J. F. (2006). Contextual diversity, not word frequency, determines word-naming and lexical decision times. *Psychological Science, 17*, 814-823.
- Andrews, S. (1997). The effect of orthographic similarity on lexical retrieval: Resolving neighborhood conflicts. *Psychonomic Bulletin & Review, 4*(4), 439-461.
- Baayen, R. H., Piepenbrock, R., & Gulikers, L. (1995). The CELEX lexical database: LDC, University of Pennsylvania, Philadelphia, PA.
- Balota, D. A., & Chumbley, J. I. (1984). Are lexical decisions a good measure of lexical access - The role of word-frequency in the neglected decision stage. *Journal of Experimental Psychology: Human Perception and Performance, 10*(3), 340-357.
- Balota, D. A., Cortese, M. J., Sergent-Marshall, S. D., Spieler, D. H., & Yap, M. J. (2004). Visual word recognition of single-syllable words. *Journal of Experimental Psychology: General, 133*(2), 283-316.
- Binder, J. R., Medler, D. A., Desai, R., Conant, L. L., & Liebenthal, E. (2005). Some neurophysiological constraints on models of word naming. *NeuroImage, 27*(3), 677-693.
- Bitan, T., & Karni, A. (2003). Alphabetical knowledge from whole words training: Effects of explicit instruction and implicit experience on learning script segmentation. *Cognitive Brain Research, 16*(3), 323-337.

- Bitan, T., & Karni, A. (2004). Procedural and declarative knowledge of word recognition and letter decoding in reading an artificial script. *Cognitive Brain Research, 19*(3), 229-243.
- Blazely, A. M., Coltheart, M., & Casey, B. J. (2005). Semantic impairment with and without surface dyslexia: Implications for models of reading. *Cognitive Neuropsychology, 22*, 695-717.
- Bodner, G. E., & Masson, M. E. J. (1997). Masked repetition priming of words and nonwords: evidence for a nonlexical basis for priming. *Journal of Memory and Language, 37*(2), 268-293.
- Bolger, D. J., Balass, M., Landen, E., & Perfetti, C. A. (2008). Context variation and definitions in learning the meanings of words: An instance-based learning approach. *Discourse Processes, 45*(2), 122-159.
- Bonin, P., Barry, C., Meot, A., & Chalard, M. (2004). The influence of age of acquisition in word reading and other tasks: A never ending story? *Journal of Memory and Language, 50*(4), 456-476.
- Bowers, J. S. (2003). An abstractionist account of masked and long-term priming. In S. Kinoshita & S. J. Lupker (Eds.), *Masked Priming: The State of the Art* (pp. 39-55). New York and Hove: Psychology Press.
- Bowers, J. S., Davis, C. J., & Hanley, D. A. (2005). Interfering neighbours: The impact of novel word learning on the identification of visually similar words. *Cognition, 97*(3), B45-B54.

- Bowers, J. S., Vigliocco, G., & Haan, R. (1998). Orthographic, phonological, and articulatory contributions to masked letter and word priming. *Journal of Experimental Psychology: Human Perception and Performance*, 24(6), 1705-1719.
- Bradley, L., & Bryant, P. E. (1979). Independence of reading and spelling in backward and normal readers. *Developmental Medicine and Child Neurology*, 21(4), 504-514.
- Bradley, L., & Bryant, P. E. (1983). Categorizing sounds and learning to read - A causal connection. *Nature*, 301(5899), 419-421.
- Brown, G. D. A., & Watson, F. L. (1987). 1st in, 1st out - Word learning age and spoken word-frequency as predictors of word familiarity and word naming latency. *Memory & Cognition*, 15(3), 208-216.
- Byrne, B. (1984). On teaching articulatory phonetics via an orthography. *Memory & Cognition*, 12(2), 181-189.
- Byrne, B., & Carroll, M. (1989). Learning artificial orthographies - Further evidence of a nonanalytic acquisition procedure. *Memory & Cognition*, 17(3), 311-317.
- Callan, A. M., Callan, D. E., & Masaki, S. (2005). When meaningless symbols become letters: Neural activity change in learning new phonograms. *NeuroImage*, 28(3), 553-562.
- Caravolas, M., Kessler, B., Hulme, C., & Snowling, M. (2005). Effects of orthographic consistency, frequency, and letter knowledge on children's vowel spelling development. *Journal of Experimental Child Psychology*, 92(4), 307-321.
- Castles, A., & Coltheart, M. (1993). Varieties of developmental dyslexia. *Cognition*, 47(2), 149-180.

- Castles, A., Davis, C., Cavalot, P., & Forster, K. (2007). Tracking the acquisition of orthographic skills in developing readers: Masked priming effects. *Journal of Experimental Child Psychology*, *97*(3), 165-182.
- Castles, A., Davis, C., & Letcher, T. (1999). Neighbourhood effects on masked form priming in developing readers. *Language and Cognitive Processes*, *14*(2), 201-224.
- Charles-Luce, J., & Luce, P. A. (1990). Similarity neighborhoods of words in young children's lexicons. *Journal of Child Language*, *17*(1), 205-215.
- Chereau, C., Gaskell, M. G., & Dumay, N. (2007). Reading spoken words: Orthographic effects in auditory priming. *Cognition*, *102*(3), 341-360.
- Chumbley, J. I., & Balota, D. A. (1984). A word's meaning affects the decision in lexical decision. *Memory & Cognition*, *12*(6), 590-606.
- Clay, F., Bowers, J. S., Davis, C. J., & Hanley, D. A. (2007). Teaching adults new words: The role of practice and consolidation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *33*(5), 970-976.
- Coltheart, M. (1978). Lexical access in simple reading tasks. In G. Underwood (Ed.), *Strategies of Information Processing*. London: Academic Press.
- Coltheart, M. (1981). The MRC psycholinguistic database. *Quarterly Journal of Experimental Psychology*, *33*(4), 497 - 505.
- Coltheart, M., Curtis, B., Atkins, P., & Haller, M. (1993). Models of reading aloud - Dual-route and parallel-distributed-processing approaches. *Psychological Review*, *100*(4), 589-608.

- Coltheart, M., Rastle, K., Perry, C., Langdon, R., & Ziegler, J. (2001). DRC: A dual route cascaded model of visual word recognition and reading aloud. *Psychological Review, 108*(1), 204-256.
- Cortese, M. J. (1998). Revisiting serial position effects in reading. *Journal of Memory and Language, 39*(4), 652-665.
- Cortese, M. J., & Fugett, A. (2004). Imageability ratings for 3,000 monosyllabic words. *Behavior Methods and Research, Instrumentation, & Computers, 36*, 84–387.
- Cortese, M. J., & Khanna, M. M. (2007). Age of acquisition predicts naming and lexical-decision performance above and beyond 22 other predictor variables: An analysis of 2,342 words. *Quarterly Journal of Experimental Psychology, 60*(8), 1072-1082.
- Cortese, M. J., & Simpson, G. B. (2000). Regularity effects in word naming: What are they? *Memory & Cognition, 28*(8), 1269-1276.
- Creel, S. C., Aslin, R. N., & Tanenhaus, M. K. (2006). Acquiring an artificial lexicon: Segment type and order information in early lexical entries. *Journal of Memory and Language, 54*(1), 1-19.
- Cutler, A., Sebastian-Galles, N., Soler-Vilageliu, O., & Van Ooijen, B. (2000). Constraints of vowels and consonants on lexical selection: Cross-linguistic comparisons. *Memory & Cognition, 28*(5), 746-755.
- Davis, C., Castles, A., & Iakovidis, E. (1998). Masked homophone and pseudohomophone priming in children and adults. *Language and Cognitive Processes, 13*(6), 625-651.

- Davis, C., Kim, J., & Forster, K. I. (2008). Being forward not backward: Lexical limits to masked priming. *Cognition*, *107*(2), 673-684.
- Dumay, N., & Gaskell, M. G. (2007). Sleep-associated changes in the mental representation of spoken words. *Psychological Science*, *18*(1), 35-39.
- Ehri, L. C. (1992). Reconceptualizing the development of sight word reading and its relationship to recoding. In P. Gough, L. C. Ehri & R. Treiman (Eds.), *Reading Acquisition* (pp. 107-143). Hillsdale, NJ: Erlbaum.
- Ellis, A. W., & Monaghan, J. (2002). Reply to Strain, Patterson, and Seidenberg (2002). *Journal of Experimental Psychology: Learning Memory and Cognition*, *28*(1), 215-220.
- Estes, K. G., Evans, J. L., Alibali, M. W., & Saffran, J. R. (2007). Can infants map meaning to newly segmented words? Statistical segmentation and word learning. *Psychological Science*, *18*(3), 254-260.
- Fodor, J. A., & Pylyshyn, Z. W. (1988). Connectionism and cognitive architecture - A critical analysis. *Cognition*, *28*(1-2), 3-71.
- Forster, K., Mohan, K., & Hector, J. (2003). The mechanics of masked priming. In S. Kinoshita & S. J. Lupker (Eds.), *Masked Priming: The State of the Art* (pp. 3-37). New York and Hove: Psychology Press.
- Forster, K. I. (1998). The pros and cons of masked priming. *Journal of Psycholinguistic Research*, *27*(2), 203-233.
- Forster, K. I. (2000). The potential for experimenter bias effects in word recognition experiments. *Memory & Cognition*, *28*(7), 1109-1115.

- Forster, K. I., & Davis, C. (1984). Repetition priming and frequency attenuation in lexical access. *Journal of Experimental Psychology: Learning Memory and Cognition*, *10*(4), 680-698.
- Frith, U. (1985). Beneath the surface of developmental dyslexia. In K. E. Patterson, J. C. Marshall & M. Coltheart (Eds.), *Surface dyslexia*. London: Erlbaum.
- Frost, S. J., Mencl, W. E., Sandak, R., Moore, D. L., Rueckl, J. G., Katz, L., et al. (2005). A functional magnetic resonance imaging study of the tradeoff between semantics and phonology in reading aloud. *Neuroreport*, *16*(6), 621-624.
- Gaskell, M. G., & Dumay, N. (2003). Lexical competition and the acquisition of novel words. *Cognition*, *89*(2), 105-132.
- Gerhand, S. (2001). Routes to reading: a report of a non-semantic reader with equivalent performance on regular and exception words. *Neuropsychologia*, *39*(13), 1473-1484.
- Gilhooley, K. J., & Logie, R. H. (1980). Age of acquisition, imagery, concreteness, familiarity, and ambiguity measures for 1944 words. *Behavior Research, Methods, & Instrumentation*, *12*, 395-427.
- Gilhooly, K., & Watson, F. (1981). Word age-of-acquisition effects: a review. *Current Psychology*, *1*(3), 269-286.
- Glushko, R. J. (1979). Organization and activation of orthographic knowledge in reading aloud. *Journal of Experimental Psychology: Human Perception and Performance*, *5*(4), 674-691.

- Gollan, T. H., Forster, K. I., & Frost, R. (1997). Translation priming with different scripts: Masked priming with cognates and noncognates in Hebrew-English bilinguals. *Journal of Experimental Psychology: Learning Memory and Cognition*, 23(5), 1122-1139.
- Graham, K. S., Hodges, J. R., & Patterson, K. (1994). The relationship between comprehension and oral reading in progressive fluent aphasia. *Neuropsychologia*, 32(3), 299-316.
- Harm, M. W., McCandliss, B. D., & Seidenberg, M. S. (2003). Modeling the successes and failures of interventions for disabled readers. *Scientific Studies of Reading*, 7(2), 155-182.
- Harm, M. W., & Seidenberg, M. S. (1999). Phonology, reading acquisition, and dyslexia: Insights from connectionist models. *Psychological Review*, 106(3), 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(3), 662-720.
- Hashimoto, R., & Sakai, K. L. (2004). Learning letters in adulthood: Direct visualization of cortical plasticity for forming a new link between orthography and phonology. *Neuron*, 42(2), 311-322.
- Hino, Y., Lupker, S. J., & Pexman, P. M. (2002). Ambiguity and synonymy effects in lexical decision, naming, and semantic categorization tasks: Interactions between orthography, phonology, and semantics. *Journal of Experimental Psychology-Learning Memory and Cognition*, 28(4), 686-713.

- Humphreys, G. W., & Evett, L. J. (1985). Are there independent lexical and nonlexical routes in word-processing? An evaluation of the dual-route theory of reading. *Behavioral and Brain Sciences*, *8*(4), 689-705.
- Jacobs, A. M., & Grainger, J. (1991). Automatic letter priming in an alphabetic decision task. *Perception & Psychophysics*, *49*(1), 43-52.
- Jacobs, A. M., & Grainger, J. (1994). Models of visual word recognition - Sampling the state-of-the-art. *Journal of Experimental Psychology: Human Perception and Performance*, *20*(6), 1311-1334.
- Jared, D. (1997). Spelling-sound consistency affects the naming of high-frequency words. *Journal of Memory and Language*, *36*(4), 505-529.
- Jared, D. (2002). Spelling-sound consistency and regularity effects in word naming. *Journal of Memory and Language*, *46*(4), 723-750.
- Jared, D., McRae, K., & Seidenberg, M. S. (1990). The basis of consistency effects in word naming. *Journal of Memory and Language*, *29*(6), 687.
- Jiang, N. (1999). Testing processing explanations for the asymmetry in masked cross-language priming. *Bilingualism: Language and Cognition*, *2*, 59-75.
- Jiang, N., & Forster, K. I. (2001). Cross-language priming asymmetries in lexical decision and episodic recognition. *Journal of Memory and Language*, *44*(1), 32-51.
- Jobard, G., Crivello, F., & Tzourio-Mazoyer, N. (2003). Evaluation of the dual route theory of reading: A metanalysis of 35 neuroimaging studies. *NeuroImage*, *20*(2), 693-712.

- Kessler, B., & Treiman, R. (2001). Relationships between sounds and letters in English monosyllables. *Journal of Memory and Language*, 44(4), 592-617.
- Kessler, B., Treiman, R., & Mullenix, J. (2003). A new look at the factors that affect oral word reading. Unpublished manuscript. .
- Knafle, J. D., & Legenza, A. (1978). Generalizability of inquiry involving artificial orthography. *American Educational Research Journal*, 15(2), 331-347.
- Kucera, H., & Francis, W. (1967). *Computational analysis of present-day American English*. Providence, RI: Brown University Press.
- Kuhl, P. K., Williams, K. A., Lacerda, F., Stevens, K. N., & Lindblom, B. (1992). Linguistic experience alters phonetic perception in infants by 6 months of age. *Science*, 255(5044), 606-608.
- Lacruz, I., & Folk, J. R. (2004). Feedforward and feedback consistency effects for high- and low-frequency words in lexical decision and naming. *Quarterly Journal of Experimental Psychology Section A: Human Experimental Psychology*, 57(7), 1261-1284.
- Laing, E., & Hulme, C. (1999). Phonological and semantic processes influence beginning readers' ability to learn to read words. *Journal of Experimental Child Psychology*, 73(3), 183-207.
- Landi, N., Perfetti, C. A., Bolger, D. J., Dunlap, S., & Foorman, B. R. (2006). The role of discourse context in developing word form representations: A paradoxical relation between reading and learning. *Journal of Experimental Child Psychology*, 94(2), 114-133.

- MacCallum, R. C., Zhang, S. B., Preacher, K. J., & Rucker, D. D. (2002). On the practice of dichotomization of quantitative variables. *Psychological Methods, 7*(1), 19-40.
- 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*(2), 202-227.
- Marks, E. A., Moates, D. R., Bond, Z. S., & Stockmal, V. (2002). Word reconstruction and consonant features in English and Spanish. *Linguistics, 40*(2), 421-438.
- Marshall, C. M., Snowling, M. J., & Bailey, P. J. (2001). Rapid auditory processing and phonological ability in normal readers and readers with dyslexia. *Journal of Speech Language and Hearing Research, 44*(4), 925-940.
- Marshall, J. C., & Newcombe, F. (1973). Patterns of paralexia: A psycholinguistic approach. *Journal of Psycholinguistic Research, 2*(3), 175-199.
- Marslen-Wilson, W. D. (1987). Functional parallelism in spoken word-recognition. *Cognition, 25*(1-2), 71-102.
- Masson, M. E. J., & Bodner, G. E. (2003). A retrospective view of masked priming: Toward a unified account of masked and long-term repetition priming. In S. Kinoshita & S. J. Lupker (Eds.), *Masked Priming: The state of the art* (pp. 57-94). New York and Hove: Psychology Press.
- Masson, M. E. J., & Isaak, M. I. (1999). Masked priming of words and nonwords in a naming task: Further evidence for a nonlexical basis for priming. *Memory & Cognition, 27*(3), 399-412.

- Masterson, J., Stuart, M., Dixon, M., & Lovejoy, S. (2003). Children's Printed Word Database. Essex, England: University of Essex.
- McClelland, J. L., & Elman, J. L. (1986). The trace model of speech-perception. *Cognitive Psychology*, *18*(1), 1-86.
- McClelland, J. L., & Rumelhart, D. E. (1981). An interactive activation model of context effects in letter perception (1). An account of basic findings. *Psychological Review*, *88*(5), 375-407.
- McKague, M., Pratt, C., & Johnston, M. B. (2001). The effect of oral vocabulary on reading visually novel words: a comparison of the dual-route-cascaded and triangle frameworks. *Cognition*, *80*(3), 231-262.
- McKay, A., Castles, A., Davis, C., & Savage, G. (2007). The impact of progressive semantic loss on reading aloud. *Cognitive Neuropsychology*, *24*(2), 162-186.
- McKay, A., Davis, C., Savage, G., & Castles, A. (2008). Semantic involvement in reading aloud: Evidence from a nonword training study. *Journal of Experimental Psychology: Learning Memory and Cognition*, *34*(6), 1495-1517.
- Mody, M., Studdert-Kennedy, M., & Brady, S. (1997). Speech perception deficits in poor readers: Auditory processing or phonological coding? *Journal of Experimental Child Psychology*, *64*(2), 199-231.
- Monaghan, J., & Ellis, A. W. (2002). What exactly interacts with spelling-sound consistency in word naming? *Journal of Experimental Psychology: Learning Memory and Cognition*, *28*(1), 183-206.

- Monsell, S. (1991). The nature and locus of word frequency effects in reading. In D. Besner & G. W. Humphreys (Eds.), *Basic processes in reading: Visual word recognition*. London: Lawrence Erlbaum Associates Inc.
- Morrison, C. M., & Ellis, A. W. (1995). Roles of word-frequency and age of acquisition in word naming and lexical decision. *Journal of Experimental Psychology: Learning Memory and Cognition*, *21*(1), 116-133.
- Morrison, C. M., Ellis, A. W., & Quinlan, P. T. (1992). Age of acquisition, not word-frequency, affects object naming, not object recognition. *Memory & Cognition*, *20*(6), 705-714.
- Morrison, C. M., & Gibbons, Z. C. (2006). Lexical determinants of semantic processing speed. *Visual Cognition*, *13*(7-8), 949-967.
- Morrison, C. M., Hirsh, K. W., Chappell, T., & Ellis, A. W. (2002). Age and age of acquisition: An evaluation of the cumulative frequency hypothesis. *European Journal of Cognitive Psychology*, *14*(4), 435 - 459.
- Morton, J. (1969). Interaction of information in word recognition. *Psychological Review*, *76*(2), 165-&.
- Nation, K., & Cocksey, J. (2009). The relationship between knowing a word and reading it aloud in children's word reading development. *Journal of Experimental Child Psychology*, *103*(3), 296-308.
- Nation, K., & Snowling, M. J. (2004). Beyond phonological skills: Broader language skills contribute to the development of reading. *Journal of Research in Reading*, *27*(4), 342-356.

- Nilsen, E., & Bourassa, D. (2008). Word-learning performance in beginning readers. *Canadian Journal of Experimental Psychology, 62*(2), 110-116.
- Norris, D. (1994). Shortlist - A connectionist model of continuous speech recognition. *Cognition, 52*(3), 189-234.
- Ouellette, G., & Fraser, J. R. (in press). What exactly is a yait anyway: The role of semantics in orthographic learning. *Journal of Experimental Child Psychology*.
- Paap, K. R., & Noel, R. W. (1991). Dual-route models of print to sound - Still a good horse race. *Psychological Research, 53*(1), 13-24.
- Paivio, A., Yuille, J. C., & Madigan, S. A. (1968). Concreteness, imagery and meaningfulness values for 925 words. *Journal of Experimental Psychology Monograph Supplement, 76*, (3, Pt 2).
- Parkin, A. J. (1982). Phonological recoding in lexical decision: Effects of spelling-to-sound regularity depend on how regularity is defined. *Memory & Cognition, 10*(1), 43-53.
- Patterson, K., & Hodges, J. R. (1992). Deterioration of word meaning: Implications for reading. *Neuropsychologia, 30*(12), 1025-1040.
- Patterson, K., Lambon-Ralph, M. A., Jefferies, E., Woollams, A., Jones, R., Hodges, J. R., et al. (2006). "Presemantic" cognition in semantic dementia: Six deficits in search of an explanation. *Journal of Cognitive Neuroscience, 18*(2), 169-183.
- Patterson, K. E. (1981). Neuropsychological approaches to the study of reading. *British Journal of Psychology, 72*(MAY), 151-174.

- Pexman, P. M., Hargreaves, I. S., Siakaluk, P. D., Bodner, G. E., & Pope, J. (2008). There are many ways to be rich: Effects of three measures of semantic richness on visual word recognition. *Psychonomic Bulletin & Review*, *15*(1), 161-167.
- Pinker, S. (1991). Rules of language. *Science*, *253*(5019), 530-535.
- Pinker, S., & Prince, A. (1988). On Language and Connectionism - Analysis of a Parallel Distributed-Processing Model of Language-Acquisition. *Cognition*, *28*(1-2), 73-193.
- Plaut, D. C. (1997). Structure and function in the lexical system: Insights from distributed models of word reading and lexical decision. *Language and Cognitive Processes*, *12*(5-6), 765-805.
- 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*(1), 56-115.
- Plunkett, K., & Juola, P. (1999). A connectionist model of English past tense and plural morphology. *Cognitive Science*, *23*(4), 463-490.
- Powell, D., Plaut, D., & Funnell, E. (2006). Does the PMSP connectionist model of single word reading learn to read in the same way as a child? *Journal of Research in Reading*, *29*(2), 229-250.
- Rack, J., Hulme, C., Snowling, M., & Wightman, J. (1994). The role of phonology in young-children learning to read words - The direct-mapping hypothesis. *Journal of Experimental Child Psychology*, *57*(1), 42-71.

- Rastle, K., & Davis, M. H. (2002). On the complexities of measuring naming. *Journal of Experimental Psychology: Human Perception and Performance*, 28(2), 307-314.
- Rastle, K., Harrington, J., & Coltheart, M. (2002). 358,534 nonwords: The ARC Nonword Database. *The Quarterly Journal of Experimental Psychology A*, 55, 1339-1362.
- Ricketts, J., Nation, K., & Bishop, D. V. M. (2007). Vocabulary is important for some, but not all reading skills. *Scientific Studies of Reading*, 11(3), 235-257.
- Rumelhart, D. E., & McClelland, J. L. (1986). On learning the past tense of English verbs. In J. L. McClelland, D. E. Rumelhart & P. R. Group (Eds.), *Parallel distributed processing: Explorations in the Microstructure of Cognition* (Vol. 2: Psychological and Biological Models, pp. 216-271). Cambridge, MA: MIT Press.
- Schwartz, M. F., Marin, O. S. M., & Saffran, E. M. (1979). Dissociations of language function in dementia: A case study. *Brain and Language*, 7(3), 277-306.
- Seidenberg, M. S., & McClelland, J. L. (1989). A distributed, developmental model of word recognition and naming. *Psychological Review*, 96(4), 523-568.
- Seidenberg, M. S., Waters, G. S., Barnes, M. A., & Tanenhaus, M. K. (1984). When does irregular spelling or pronunciation influence word recognition. *Journal of Verbal Learning and Verbal Behavior*, 23(3), 383-404.
- Seymour, P. H. K., Aro, M., Erskine, J. M., Wimmer, H., Leybaert, J., Elbro, C., et al. (2003). Foundation literacy acquisition in European orthographies. *British Journal of Psychology*, 94, 143-174.

- Snowling, M. J., Gallagher, A., & Frith, U. (2003). Family risk of dyslexia is continuous: Individual differences in the precursors of reading skill. *Child Development, 74*(2), 358-373.
- Stager, C. L., & Werker, J. F. (1997). Infants listen for more phonetic detail in speech perception than in word-learning tasks. *Nature, 388*(6640), 381-382.
- Stone, G. O., Vanhoy, M., & VanOrden, G. C. (1997). Perception is a two-way street: Feedforward and feedback phonology in visual word recognition. *Journal of Memory and Language, 36*(3), 337-359.
- Strain, E., Patterson, K., & Seidenberg, M. S. (1995). Semantic effects in single-word naming. *Journal of Experimental Psychology: Learning Memory and Cognition, 21*(5), 1140-1154.
- Strain, E., Patterson, K., & Seidenberg, M. S. (2002). Theories of word naming interact with spelling-sound consistency. *Journal of Experimental Psychology: Learning Memory and Cognition, 28*(1), 207-214.
- Stuart, M., Masterson, J., Dixon, M., & Quinlan, P. (1999). Inferring sublexical correspondences from sight vocabulary: Evidence from 6-and 7-year-olds. *Quarterly Journal of Experimental Psychology Section A: Human Experimental Psychology, 52*(2), 353-366.
- Toglia, M. P., & Battig, W. F. (1978). *Handbook of Semantic Word Norms*. Hillsdale, NJ: Erlbaum.
- Treiman, R., Goswami, U., & Bruck, M. (1990). Not all nonwords are alike - Implications for reading development and theory. *Memory & Cognition, 18*(6), 559-567.

- Treiman, R., & Kessler, B. (2006). Spelling as statistical learning: Using consonantal context to spell vowels. *Journal of Educational Psychology, 98*(3), 642-652.
- Treiman, R., Kessler, B., & Bick, S. (2003). Influence of consonantal context on the pronunciation of vowels: A comparison of human readers and computational models. *Cognition, 88*(1), 49-78.
- Treiman, R., Kessler, B., Zevin, J. D., Bick, S., & Davis, M. (2006). Influence of consonantal context on the reading of vowels: Evidence from children. *Journal of Experimental Child Psychology, 93*(1), 1-24.
- Treiman, R., Mullennix, J., Bijeljacabic, R., & Richmond-Welty, E. D. (1995). The special role of rimes in the description, use, and acquisition of English orthography. *Journal of Experimental Psychology: General, 124*(2), 107-136.
- Trudeau, J. (2006). *Semantic contributions to word naming with artificial lexicons*. Unpublished doctoral dissertation, University of Connecticut - Connecticut.
- Unsworth, S. J., & Pexman, P. M. (2003). The impact of reader skill on phonological processing in visual word recognition. *Quarterly Journal of Experimental Psychology Section A: Human Experimental Psychology, 56*(1), 63-81.
- Van Ooijen, B. (1996). Vowel mutability and lexical selection in English: Evidence from a word reconstruction task. *Memory & Cognition, 24*(5), 573-583.
- Ventura, P., Morais, J., Pattamadilok, C., & Kolinsky, R. (2004). The locus of the orthographic consistency effect in auditory word recognition. *Language and Cognitive Processes, 19*(1), 57-95.

- Waters, G. S., & Seidenberg, M. S. (1985). Spelling-sound effects in reading: Time-course and decision criteria. *Memory & Cognition*, *13*(6), 557-572.
- Waters, G. S., Seidenberg, M. S., & Bruck, M. (1984). Childrens and adults use of spelling-sound information in 3 reading tasks. *Memory & Cognition*, *12*(3), 293-305.
- Weekes, B. S., Castles, A. E., & Davies, R. A. (2006). Effects of consistency and age of acquisition on reading and spelling among developing readers. *Reading and Writing*, *19*(2), 133-169.
- Werker, J. F., Cohen, L. B., Lloyd, V. L., Casasola, M., & Stager, C. L. (1998). Acquisition of word-object associations by 14-month-old infants. *Developmental Psychology*, *34*(6), 1289-1309.
- Werker, J. F., Fennell, C. T., Corcoran, K. M., & Stager, C. L. (2002). Infants' ability to learn phonetically similar words: Effects of age and vocabulary size. *Infancy*, *3*(1), 1-30.
- Werker, J. F., & Tees, R. C. (1984). Cross-language speech-perception - Evidence for perceptual reorganization during the 1st year of life. *Infant Behavior & Development*, *7*(1), 49-63.
- Whaley, C. P. (1978). Word-nonword classification time. *Journal of Verbal Learning and Verbal Behavior*, *17*(2), 143-154.
- Wonnacott, E., Newport, E. L., & Tanenhaus, M. K. (2008). Acquiring and processing verb argument structure: Distributional learning in a miniature language. *Cognitive Psychology*, *56*(3), 165-209.

- Woollams, A. M. (2005). Imageability and ambiguity effects in speeded naming: Convergence and divergence. *Journal of Experimental Psychology: Learning Memory and Cognition*, 31(5), 878-890.
- Woollams, A. M., Lambon-Ralph, M. A., Plaut, D. C., & Patterson, K. (2007). SD-squared: On the association between semantic dementia and surface dyslexia. *Psychological Review*, 114(2), 316-339.
- Zevin, J. D., & Seidenberg, M. S. (2004). Age-of-acquisition effects in reading aloud - Tests of cumulative frequency and frequency trajectory. *Memory & Cognition*, 32(1), 31-38.
- Ziegler, J. C., & Ferrand, L. (1998). Orthography shapes the perception of speech: The consistency effect in auditory word recognition. *Psychonomic Bulletin & Review*, 5(4), 683-689.
- Zorzi, M., Houghton, G., & Butterworth, B. (1998). Two routes or one in reading aloud? A connectionist dual-process model. *Journal of Experimental Psychology: Human Perception and Performance*, 24(4), 1131-1161.

APPENDICES

Appendix 2.A. Pronunciations of Training Items used in Experiment 1

Orthography	Consistent pronunciation	Inconsistent pronunciation
MAVE	meiv	mæv
LUT	lʌt	lʊt
SPALL	spɔl	spæl
RINT	rɪnt	raɪnt
TROSS	trɒs	trəʊs
TOUND	taʊnd	tʊnd
NOWL	naʊl	nəʊl
NAFT	nɔft	nɒft
KROUGHT	krɒt	kraʊt
STOME	stəʊm	stʌm
VORD	vɒd	vɜd
KAID	keɪd	kɛd
FRASTE	freɪst	frast
RONK	rɒŋk	rʌŋk
TROAD	trəʊd	trɒd
GLORM	glɔm	glɜm
FLEIN	flem	flaɪn
FREW	fru	frəʊ
NEAK	nik	neɪk
PIVE	pɑɪv	pɪv
ZOOK	zʊk	zʊk
DROSE	drəʊz	druz
SCORSE	skɒs	skɜs
CHEY	tʃeɪ	tʃi

Appendix 2.B. Definitions used in Experiment 1

High imageability definitions

An antique luxury car

A wig made from real human hair

A kind of vest with several large pockets

The device used to milk dairy cows

Open-topped cargo container for bulk shipping

The blade of an ice-skate

Child's toy made from a large ring and several small rings

A double-barrelled suction cup dart gun

Metal tool for cutting plastic cables

A knife with a hooked blade

A medieval stringed instrument

A bag used for carrying dental instruments

South American rodent with a long tail and large ears

Sausage made with fish meat

A textured metal hammer for tenderizing meat

A mask of the sort worn by super-heroes

A kind of duck with a green head and white chest

A combination camera and microphone used with computers

Jewellery made by coating flowers with gold

Oversized clown shoes

An adaptor used to connect several audio speakers together

A plant with five-leaved blue flowers

An old-fashioned rotary dial phone

Appendix 2.B. continued.

Low imageability definitions

Brightness of a colour

Built according to a set of strict rules or specifications

Using a tool for a task it wasn't intended for

Doing things out of obligation

Amount of energy produced by a machine

Taking blame when actually innocent

Something that is larger than it should be

Seeing similarities between facts

A false advertisement

Someone who doesn't return calls

A portion of something that is representative of the whole

The status of being protected from dismissal

Being a neutral judge in an argument

A joke at someone else's expense

Watching someone else make a mistake

The ability to tell the time without a clock

Being short-tempered and unreasonable

To express disdain or reproach

A false name used during identity theft

When something is very important but still boring

The stress of having no money





A person who you pretend to like but really don't

Lacking in knowledge

Appendix 3.A. Novel Characters and Phoneme Mappings used in Experiment 2

Character	Phoneme Mapping 1	Phoneme Mapping 2
Λ	/u/ or /ʌ/	/aɪ/ or /ɪ/
≠	/b/	/f/
Ʒ	/p/	/m/
ℓ	/i/ or /ɛ/	/əʊ/ or /ɒ/
⊗	/əʊ/ or /ɒ/	/i/ or /ɛ/
Ɔ	/d/	/b/
Ɔ	/s/	/n/
ε	/m/	/z/
⊙	/f/	/g/
≡	/aɪ/ or /ɪ/	/u/ or /ʌ/
▭	/v/	/s/
∇	/n/	/v/
Ɔ	/z/	/p/
Ɔ	/g/	/d/
Ɔ	/t/	/k/
⊙	/k/	/t/

Appendix 3.B. Frequency x Consistency manipulation used in Experiment 2 for stimulus set 1, character-phoneme mapping 1

Character	Pronunciation	Consistency	Onset	Frequency
	/u/	Consistent	any	8 items
	/aɪ/	Consistent	any	4 items
	/əʊ/	Inconsistent-conditioned	/z/ character	4 items
	/ɒ/	Inconsistent-unconditioned	any except /z/ character	8 items
	/i/	Inconsistent-conditioned	/v/ character	8 items
	/ɛ/	Inconsistent-unconditioned	any except /v/ character	4 items

Appendix 3.C. Pronunciations of Training and Generalization items in Experiment 2

Table 3.C1

Training Item Pronunciations

High Frequency			Low Frequency		
Cons	Incons-cond	Incons-uncond	Cons	Incons-cond	Incons-uncond
<u>Stimulus Set 1</u>					
buy, dus, fun, tup, kuf, nut, pug	vid, vif, vig, vik, vim, vip, vis, viv	bɒz, gɒm, mɒn, pɒs, dɒz, vɒb, fɒt, nɒf	kaid, saiv, zaik, gam	zəʊb, zəʊm, zəʊt, zəʊv	tez, mep, zek, sev
<u>Stimulus Set 2</u>					
big, diz, kit, pif, sib, tip, vin, zis	vud, vuf, vug, vuk, vum, vup, vuv, vuz	div, fik, gim, kiz, mip, nif, pib, vin	bəʊs, fəʊd, gəʊk, təʊm	zaig, zam, zait, zaiv	mɒb, nɒs, sɒt, zɒd

Table 3.C2

Generalization Item Pronunciations

High Frequency			Low Frequency		
Cons	Incons-cond	Incons-uncond	Cons	Incons-cond	Incons-uncond
<u>Stimulus Set 1</u>					
fub, mup	viz, vin	vɒn, pɒf	baɪm, zaɪg	zəʊd, zəʊs	gem, nev
<u>Stimulus Set 2</u>					
mig, siv	vub, vus	pɪd, vɪf	dəʊn, zəʊg	zaɪb, zaɪs	fɒp, gɒz

Appendix 3.D. Results of analyses of variance on Old-New Decision performance in
Experiment 2

Effect	Subjects analysis statistics	Items analysis statistics
<u>Accuracy</u>		
Frequency	$F_p < 1$	$F_i < 1$
Consistency	$F_p(2, 22) = 2.00, p = .16, \eta_p^2 = .15$	$F_i(2, 18) = 1.92, p = .18, \eta_p^2 = .18$
Freq x Cons	$F_p < 1$	$F_i < 1$
<u>RTs</u>		
Frequency	$F_p < 1$	$F_i < 1$
Consistency	$F_p < 1$	$F_i < 1$
Freq x Cons	$F_p(2, 22) = 1.22, ns$	$F_i < 1$

Appendix 4.A. Untrained items for Generalization and Old-New Decision on Day 2 -

Experiment 3

High Frequency			Low Frequency		
Cons	Incons-cond	Incons-uncond	Cons	Incons-cond	Incons-uncond
<u>Stimulus Set 1</u>					
kus, vun	vib, vit	fɒd, sɒm	daɪt, maɪz	zəʊk, zəʊp	bɛf, nɛg
<u>Stimulus Set 2</u>					
dis, vim	vun, vut	bɪp, mɪz	fəʊv, zəʊb	zaɪd, zaɪk	kʌg, sʌf

Appendix 4.B. Results of analysis of variance on Accuracy in Old-New Decision in
Experiment 3

Effect	Subjects analysis statistics	Items analysis statistics
Day	$F_p(1, 15) = 4.35, p = .06, \eta_p^2 = .23$	$F_i(1, 36) = 2.26, p = .14, \eta_p^2 = .06$
Frequency	$F_p(1, 15) = 1.65, ns$	$F_i(1, 36) = 1.18, ns$
Consistency	$F_p < 1$	$F_i < 1$
Freq x Cons	$F_p(2, 30) = 1.13, ns$	$F_i < 1$
Freq x Day	$F_p(1, 15) = 3.27, p = .09, \eta_p^2 = .18$	$F_i(1, 36) = 1.77, p = .19, \eta_p^2 = .05$
Cons x Day	$F_p < 1$	$F_i < 1$
Freq x Cons x Day	$F_p(2, 30) = 2.46, p = .10, \eta_p^2 = .14$	$F_i > 1$

Appendix 5.A. Examples of Novel Objects used in Experiment 4



Appendix 5.B. Control Primes used in Old-New Decision in Experiment 4

Trained item target	Control prime	Untrained item distractor	Control prime
<u>Stimulus Set 1</u>			
fɒt	fɑɪg	bɑɪm	bɛp
mɛp	mʊb	fʊb	fɒs
nɒf	nʊf	gɛm	gɒf
pʊg	pɒv	mʊp	mɑɪd
sɑɪv	sɛm	nɛv	nɑɪg
sɛv	sɒt	pɒf	pɛb
tʊp	tɑɪf	vɪn	vɒd
vɪg	vɑɪf	vɪz	vɒm
vɪs	vɒk	vɒn	vʊg
zəʊb	zɑɪn	zəʊd	zɛb
zəʊm	zɛf	zəʊs	zɛm
zɑɪk	zʊn	zɑɪg	zɛs
<u>Stimulus Set 2</u>			
bəʊs	bʌv	dəʊn	dɪf
bɪg	bʌp	fʌp	fɪm
dɪz	dəʊk	gʌz	ɡɪt
gəʊk	ɡɪp	mɪg	mʌn
mʌb	mɪg	pɪd	pəʊs
nɪf	nʌv	sɪv	sɪz
sʌt	sɪg	vɪf	vəʊm
vɪn	vɪz	vʊb	vɪf
vʊd	vɪb	vʊs	vɪg
vʊm	vəʊn	zəʊg	zɑɪm
zɑɪn	zʌb	zɑɪb	zʌf
zɑɪt	zʌm	zɑɪs	zʌg

Appendix 6.A. Definitions used in Experiment 5

The fold of skin hanging from the neck of cattle

A wooden case used for storing canon-balls

A cape worn by a bull fighter

An intoxicating liquor made from sweet potatoes

A wooden beam across the opening of a fireplace

A small wooden truck which runs on rails

A dangerous stretch of water near a reef

An assistant to a magician or scholar

A Mexican shelter made of palm leaves

A leather bottle used for storing oils

An instrument that measures time by the flow of water

Material which accumulates at the bottom of a steep slope

A long horn used by herdsmen

A workman who unloads fish from trawlers

The innermost part of a temple

The screen that separates the choir from the body of a church

The rosy light of the sun seen on mountains

A black hole formed by the death of a star

Appendix 6.B. Effects of Frequency and Consistency on Performance at the Beginning and End of Training in Experiment 5, and in comparisons with Experiments 3 and 4.

Table 6.B1

Beginning of training

Effect	Subjects analysis statistics	Items analysis statistics
Semantics vs. Lexical Phonology (Experiment 5)		
Frequency	$F_p(1, 31) = 23.97, p < .001, \eta_p^2 = .44$	$F_i(1, 66) = 16.97, p < .001, \eta_p^2 = .21$
Consistency	$F_p(2, 62) = 3.81, p < .05, \eta_p^2 = .11$	$F_i(2, 66) = 3.83, p < .05, \eta_p^2 = .10$
Freq x Cons	$F_p(2, 62) = 1.64, ns$	$F_i(2, 66) = 1.15, ns$
Semantics (Exp 5) vs. Lex Phon (Exp 4) vs. None (Exp 3)		
Frequency	$F_p(1, 31) = 36.75, p < .001, \eta_p^2 = .38$	$F_i(1, 66) = 26.78, p < .001, \eta_p^2 = .29$
Consistency	$F_p(2, 122) = 11.99, p < .001, \eta_p^2 = .16$	$F_i(2, 66) = 10.67, p < .001, \eta_p^2 = .24$
Freq x Cons	$F_p(2, 122) = 1.39, ns$	$F_i(2, 66) = 1.06, ns$
Lex Phon (Exp 5) vs. Lex Phon (Exp 4) vs. None (Exp 3)		
Frequency	$F_p(1, 61) = 45.91, p < .001, \eta_p^2 = .43$	$F_i(1, 66) = 31.95, p < .001, \eta_p^2 = .33$
Consistency	$F_p(2, 122) = 10.07, p < .001, \eta_p^2 = .14$	$F_i(2, 66) = 9.20, p < .001, \eta_p^2 = .22$
Freq x Cons	$F_p(2, 122) = 1.59, ns$	$F_i < 1$

Appendix 6.B. continued

Table 6.B2

End of training

Effect	Subjects analysis statistics	Items analysis statistics
Semantics vs. Lexical Phonology (Experiment 5)		
Frequency	$F_p(1, 31) = 43.72, p < .001, \eta_p^2 = .59$	$F_i(1, 66) = 41.96, p < .001, \eta_p^2 = .39$
Consistency	$F_p(2, 62) = 7.35, p = .001, \eta_p^2 = .19$	$F_i(2, 66) = 9.28, p < .001, \eta_p^2 = .22$
Freq x Cons	$F_p(2, 62) = 4.76, p = .01, \eta_p^2 = .13$	$F_i(2, 66) = 4.74, p = .01, \eta_p^2 = .13$
Semantics (Exp 5) vs. Lex Phon (Exp 4) vs. None (Exp 3)		
Frequency	$F_p(1, 61) = 49.78, p < .001, \eta_p^2 = .45$	$F_i(2, 66) = 42.22, p < .001, \eta_p^2 = .39$
Consistency	$F_p(2, 122) = 19.40, p < .001, \eta_p^2 = .24$	$F_i(2, 66) = 17.42, p < .001, \eta_p^2 = .35$
Freq x Cons	$F_p(2, 122) = 3.87, p < .05, \eta_p^2 = .06$	$F_i(2, 66) = 3.72, p < .05, \eta_p^2 = .10$
Lex Phon (Exp 5) vs. Lex Phon (Exp 4) vs. None (Exp 3)		
Frequency	$F_p(1, 61) = 68.45, p < .001, \eta_p^2 = .53$	$F_i(1, 66) = 50.28, p < .001, \eta_p^2 = .43$
Consistency	$F_p(2, 122) = 16.80, p < .001, \eta_p^2 = .22$	$F_i(2, 66) = 14.72, p < .001, \eta_p^2 = .31$
Freq x Cons	$F_p(2, 122) = 6.04, p < .01, \eta_p^2 = .09$	$F_i(2, 66) = 5.39, p < .01, \eta_p^2 = .14$