What's in a Story?

Children's Learning of New Written Words

via Reading Experience

Niina Tamura

St John's College, University of Oxford

Thesis submitted for the degree of Doctor of Philosophy

Michaelmas Term, 2015

Table of Contents

Acknowledgements

First, I would like to thank my supervisor, Kate Nation, for her unwavering support and guidance throughout my time in Oxford. Kate, I can't thank you enough for taking me on as your student, and for everything that followed. In addition, I would like to thank Anne Castles, Sachiko Kinoshita, and Bob McMurray for their guidance along the way. Special thanks to past and present members of LCD, in particular Nick Cooper and Bram Vandekerckhove for their patient support in all statistical matters, and Becca Wedge-Roberts for help with data collection. I was fortunate to receive the support of many different people over the course of my time in Oxford, all of whom I would like to thank, with special mention to Elsa Bell, and my college advisor, Dorothy Bishop. For financial support, I would like to thank the Economic and Social Research Council and St John's College. I also wish to thank the children, parents, teachers and staff who made this research possible: Dr South, Middleton Cheney, Riverside, St. Ebbe's, and St. Patrick's. Finally, I would like to thank my parents, my friends, and Joel, for their unending support and encouragement throughout.

Abstract

What's in a Story?
Children's Learning of New Written Words via Reading Experience

Niina Tamura

St John's College, University of Oxford

Thesis submitted for the degree of Doctor of Philosophy

Over the course of reading development, children are able to recognise an increasingly large number of words by sight, a process termed orthographic learning, as they move from effortful decoding to fast, efficient word recognition. Phonological decoding has been shown to account for some, but not all variance in orthographic learning. While it is clear that reading experience plays a crucial role, the mechanisms underlying orthographic learning require further investigation. The research presented in this thesis had two overarching objectives: first, to establish a novel, sensitive measure of orthographic learning, using the prime-lexicality effect in lexical decision to index the lexicalisation of novel words, and second, to investigate the effects of specific parameters of reading experience on the learning of novel words. In particular, the experiments focussed on incidental, non-directed learning of unfamiliar English words via story reading. Learning was assessed in terms of knowledge of the novel word spellings and meanings themselves, and their engagement with existing words in the lexicon, indexed by the prime-lexicality effect. Five experiments manipulated directed vs. incidental learning, number of exposures, the availability of semantics, and the variability of the story contexts. Across experiments, children showed good learning of the novel words, and a prime-lexicality effect emerged, validating this novel measure. Number of exposures affected lexicalisation of the novel words, as a prime-lexicality effect was evident after twelve exposures, but not after four. Performance was otherwise unaffected by the experimental manipulations, revealing robust orthographic learning in children aged 9 to 11.

Chapter 1:

General Introduction

Being able to read words correctly and efficiently is a fundamental component of reading. Much research has focussed on how beginning readers discover the alphabetic principle and develop the ability to decode words. Comparatively less is known about how children move from effortful decoding to skilled, efficient word recognition (Nation & Castles, in press). However, without a doubt, reading experience plays a crucial role in this process. Not only is print exposure related to reading development in general (Mol & Bus, 2011), at an item-level, experiencing unfamiliar words in meaningful context offers a fruitful learning opportunity (e.g. Nagy, Herman, & Anderson, 1985; Share, 1995). The research reported in this thesis investigated children's learning of novel written words via reading experience. Crucially, this required a sensitive measure of learning, which was developed as part of this thesis. This chapter provides a broad overview of the literature concerning the research questions investigated by this thesis. More detailed discussion of specific areas of research is provided in the introduction to each individual chapter.

1.1.Orthographic Learning in the Context of Reading Development

Learning to read is a complex process. Over the course of reading development, children progress from knowledge of isolated words and individual letters to being able to comprehend extended texts. The present research focuses on children's ability to read words, in particular the development of skilled, automatic word recognition ('sight word reading'). Ehri (2005, 2014) has characterised this development in four phases. Throughout, learning to read words by sight is described as a process of forming connections or mappings between visual units (e.g. letters) and sounds (e.g. phonemes). This process depends crucially on alphabetic knowledge, and as this evolves, the predominant type of connections changes, characterizing the four phases. In the *pre-alphabetic phase,* children are essentially pre-reading and have no alphabetic knowledge. They may be able to read a few words based on visual or contextual cues, such as logos on common items, or their own name. In the *partial alphabetic phase*, children begin to acquire knowledge of letter names or sounds, which allows them to connect some of the salient letters in words to sounds, e.g. first and final letters. However, they are not yet able to fully decode words into all their letters and corresponding phonemes, and thus may confuse words with similar spellings. In the *full alphabetic phase*, children's knowledge of grapheme-phoneme correspondences has evolved to allow them to fully decode words. Children begin to retain spellings of specific words in memory, which they can then recognise automatically on subsequent encounters. However, the sum of words they are able to recognise by sight (termed *sight word memory* or *print lexicon*) is still relatively limited. In the *consolidated alphabetic phase*, children have acquired knowledge of common spelling patterns and their corresponding pronunciations, allowing them to recognise words based on larger chunks (e.g. syllables or morphemes) than individual letters. This knowledge of spelling patterns also allows them to read words by analogy. Importantly, these phases are not suggested to constitute distinct stages. At any point in development, children use more than one type of connection, e.g. in the full alphabetic stage, they may decode most words based on grapheme-phoneme correspondences, but also recognise some by sight. The four phases are characterised based on the type of connection used most often, and have been argued to constitute overlapping waves rather than discreet stages (Ehri, 2014).

Ehri's (2005, 2014) phases offer a useful description of the reading profiles children show as they progress towards skilled reading. Subsequent research has investigated the underlying cognitive skills that likely underpin children's reading abilities, e.g. by measuring potential predictors prior to formal reading instruction and relating them to later reading outcomes (reviewed in Hulme & Snowling, 2014). Much of this research has focused on predicting relatively early reading skills, i.e. children's ability to map letters on to sounds, which allows them to phonologically decode words. There is now good evidence that decoding ability is tightly bound with spoken language skills. Longitudinal studies assessing oral language skills before or at the start of reading instruction have shown that phoneme awareness (the ability to identify and manipulate phonemes in spoken language), letter knowledge (knowledge of letter names and sounds), and rapid automatised naming (speeded retrieval of the names of visual stimuli, e.g. pictures) predict later decoding skill. Further evidence for a causal relationship between phoneme awareness, as well as letter knowledge, and reading outcomes comes from training studies (reviewed in Hulme & Snowling, 2014).

The present research focuses on children who have mastered these early decoding skills, and asks the question of how they increase their print lexicon. In this respect, the term *orthographic learning* is used to describe the process of moving from effortful decoding of a word to skilled, automatic recognition (Castles & Nation, 2006). There is both theoretical and empirical support for the notion that word recognition is an important component of successful reading comprehension, arguably the ultimate goal of reading. The Simple View of Reading suggests that for successful reading comprehension, the reader needs to be proficient in two skills – listening comprehension and visual word recognition/decoding (Gough & Tunmer, 1986). At the word level, the lexical quality hypothesis suggests that accurate and efficient reading comprehension depends on adequate lexical quality (Perfetti & Hart, 2002). The lexical quality of a given word is determined by phonological, orthographic and meaning components, which together form word identity. Empirical support for the relative importance of the orthographic component was provided by a cross-sectional study of German primary school children. Richter, Isberner, Naumann, and Neeb (2013) showed that while semantic skills partially mediated the effect of orthographic and phonological skills on reading comprehension, there was also a direct effect of orthographic skills on reading comprehension that was much stronger than that of phonological skills. There is no doubt that typically developing children generally progress toward skilled word reading as described by Ehri (2005, 2014), but what are the underlying mechanisms? As in research into the foundations of phonological decoding, predictors of orthographic learning and visual word recognition in children have been investigated, but arguably with more limited success (reviewed in Castles & Nation, 2006).

One predictor is children's rudimentary reading ability itself, i.e. alphabetic skills, decoding ability, and phonological skills. This consideration will be discussed in more detail below when considering mechanisms of orthographic learning. Second, orthographic processing skills, e.g. knowledge of orthographic regularities, have been suggested to explain variance in orthographic learning. However, Castles and Nation (2006) identified key issues with this line of research. First, there is a problem with circularity as most measures of orthographic processing are difficult to distinguish from the predicted outcome itself, i.e. word recognition. Second, one can question whether orthographic processing skills are an outcome of successful orthographic learning, rather than an underlying cause.

Much of the research described above focuses on child-level predictors of word reading, that is, the cognitive skills which might underpin successful reading. A different angle, taken for the present research, is to focus on the item level, i.e. the question of how written words become skillfully recognised. While children bring general knowledge of their writing system to the task, experience of specific words is what allows them to be read by sight. The following focuses on proposed mechanisms of how item-specific orthographic knowledge is acquired.

1.2 The Self-teaching Hypothesis

The self-teaching hypothesis postulates that during independent reading, phonological (letter-to-sound) decoding serves as the primary mechanism for children to acquire word-specific orthographic representations (Share, 1995). When children encounter a written word they have not seen before, they can use their decoding skills to arrive at its pronunciation. This allows them to acquire word-specific orthographic knowledge, which they can use on subsequent encounters to allow fast, efficient recognition. In the classic paradigm, children read aloud short texts containing pseudowords (e.g. *yait* in English language experiments). Orthographic learning is subsequently measured using tasks such as orthographic choice, naming, and spelling, discussed in more detail later on. Using this paradigm, it has been shown that very few exposures to a word (in some cases just one) can suffice to acquire some orthographic knowledge in Hebrew (Share, 2004), Dutch (de Jong & Share, 2007), and English (Cunningham, Perry, Stanovich, & Share, 2002; Nation, Angell, & Castles, 2007). This provides evidence that orthographic learning does not require direct instruction, i.e. can take place incidentally.

In support of the emphasis on phonological decoding, positive correlations between target word decoding and subsequent orthographic learning have been reported (e.g. r = .52, Cunningham et al., 2002). However, these correlations are not perfect, suggesting that there are other factors not accounted for. At an item-level, Nation et al. (2007) also reported several cases of decoding success but learning failure, and vice versa. The self-teaching hypothesis does acknowledge developmental changes as the body of acquired orthographic knowledge increases (Share, 1995), which might account for some of the unexplained variance. As readers progress, orthographic regularities and constraints are extrapolated from reading experience, informing the process of phonological decoding.

In line with this, these orthographic processing skills have been suggested to play a role in orthographic learning itself, rather than merely result from it. Differences in orthographic processing skills have been found to predict orthographic learning success (e.g. Wang, Nickels, Nation, & Castles, 2013), but the concerns outlined above with regard to predicting children's skilled reading also apply to predicting learning success, as measures of orthographic processing skill and orthographic learning are difficult to separate (Castles & Nation, 2006; Nation & Castles, in press). For example, Wang et al. (2013) assessed orthographic knowledge using reading accuracy on irregular words in order to predict orthographic learning. Children's ability to correctly read irregular words is likely determined by earlier orthographic learning. As such, demonstrating that earlier orthographic learning success predicts subsequent orthographic learning success adds little explanatory power with regard to the underlying mechanisms. Together, the above considerations suggest that the self-teaching hypothesis is currently underspecified. While it broadly describes how knowledge is added to the visual word recognition system, it does not offer obvious predictions regarding factors beyond decoding skills that may or may not influence orthographic learning. In this thesis, the focus is shifted away from children's underlying cognitive skills to their specific experience of written words.

1.3 Orthographic Learning via Reading Experience

The importance of children's reading experience and practice for learning to read is well documented. Broadly speaking, print exposure is positively associated with the development of reading skill. It has been suggested that this relationship likely reflects reciprocal causation, where print exposure enables children to improve their skills and knowledge, and children with better reading skills also choose to engage in reading activities more than children with poorer reading skills (Mol & Bus, 2011). In this thesis, the focus is on the former relationship, in particular the role of reading experience in driving learning about words. Although children come to the task of learning to read with knowledge of spoken words, children also frequently encounter entirely novel words during reading, i.e. words that are unfamiliar in spoken language (Nagy et al., 1985). For instance, children's books have been estimated to contain three times as many low-frequency words than TV shows or conversations between adults and children (Hayes & Ahrens, 1988). Encountering unfamiliar words during reading offers an opportunity for incidental learning, i.e. learning without direct instruction (Joseph, Wonnacott, Forbes, & Nation, 2014; Nagy et al., 1985). Incidental learning is discussed in more detail in Chapter 2. Regarding the types of materials children read, this thesis largely focussed on story reading. Fiction has been hypothesised to be particularly conducive to learning about new words, potentially due to effects of motivation and emotional engagement (Mol & Bus, 2011).

The research reported here primarily investigated learning of novel word forms, with a secondary focus on learning about meanings. Unequivocally, reading experience has to be the gateway to high-quality orthographic knowledge of words. As Perfetti and Hart (2002, p. 192) suggest, 'The genesis of the number of quality word representations is a matter of experience that produces relevant knowledge'. A word that has not been experienced at all by a child by definition has zero lexical quality. Exposure via reading experience offers an opportunity for word learning, and thus the development of lexical quality. Specifically, orthographic learning can be considered the process that provides the orthographic component of lexical quality. At its most basic, the framework for orthographic learning in the present research can thus be summed up as follows:

Reading experience → *exposure* → *orthographic learning* → *lexical quality*

For any given word, the above process can be repeated with recurring exposures.

The importance of experience is also an integral part of connectionist models of reading aloud and word recognition (e.g. Harm & Seidenberg, 2004; Seidenberg & McClelland, 1989.; Ziegler, Perry, & Zorzi, 2014). Details of specific models are beyond the scope of this thesis, but generally speaking, skilled model performance is the result of learning via exposures to written word input, upon which the model generates output (e.g. a phonological form or a meaning) via connections. Each exposure offers an opportunity for learning. Models vary in the extent to which learning is implemented in a way that is developmentally plausible; for example, some require an unrealistically high number of training trials, as well as supervised learning (Harm & Seidenberg, 2004; see Castles & Nation, 2006, for further discussion). Addressing these concerns, Ziegler et al. (2014) proposed a connectionist dual processing (CDP) model, which demonstrates unsupervised learning akin to self-teaching after a relatively few exposures. A further feature of models of visual word recognition is the notion of competition effects (see Norris, 2013, for discussion of specific models). Lexical competition arises between orthographically similar words, i.e. words which overlap in letters. The precise origin of competition effects varies between models depending on their architecture and implementation. For example, in the classic interactive activation model, competition arises due to inhibitory links between words nodes (McClelland & Rumelhart, 1981; Rumelhart & McClelland, 1982). For the purpose of the present discussion, what is most relevant is the link between experience, orthographic learning, and the emergence of lexical competition as the orthographic lexicon increases.

Reading experience not only allows orthographic learning of specific words, but also has consequences for a child's visual word recognition system as a whole, as described by the lexical tuning hypothesis (Castles, Davis, Cavalot, & Forster, 2007). This hypothesis explains the developmental change resulting from orthographic learning and the associated increase in orthographic knowledge ('lexical density', Castles et al. 2007), akin to the emergence of lexical competition. As children acquire a wider written vocabulary, the word recognition system is said to evolve from being fairly broadly tuned to becoming more finely tuned. When the number of entries in the orthographic lexicon is small, there are likely to be fewer highly similar, competing items. Thus the system can afford to be broadly tuned while still being able to accurately distinguish and recognise words. As written vocabulary increases, the greater number of similar items also requires a system that is more stringent to maintain accuracy, i.e. more finely tuned. This suggests that developmental change is driven by the intrinsic nature of the word recognition system, in particular knowledge acquired via experience. As readers improve, they are expected to show evidence of fine-grained distinctions not simply because the system has become able to make them, but also because it has become advantageous to be more stringent in terms of efficiency. There is likely to be a bidirectional relationship, as the state of the visual word recognition system in turn could have implications for orthographic learning.

The lexical tuning hypothesis is supported by a number of studies using masked priming in word recognition tasks. For instance, Castles et al. (2007) investigated letter substitution priming, using nonword primes with one letter different from target words. They compared Grade-3 children to skilled adults on a primed visual lexical decision task using target words with high-density orthographic neighbourhoods and high in frequency. Adults did not show evidence of form priming, as expected given that the target words required precise discrimination. In contrast to this, Grade-3 children did show evidence of one-letter-different form priming. When children were retested in Grade 5, the adult pattern emerged. This shift was argued to be evidence for a lexical tuning process.

1.3.1 Specific parameters of reading experience and their influence on learning

Having established that reading experience is crucial, the logical next step is to question which of its parameters are important for orthographic learning, as children's reading experience is rich and varied. There are a number of candidate factors, some of which have been investigated by existing research. Areas chosen for investigation in this research are summarised below, with more detailed discussion in the relevant chapters.

Incidental vs. directed exposure. In the literature on vocabulary acquisition, incidental and directed learning are commonly contrasted (Swanborn & de Glopper, 1999). Incidental learning has been defined as 'the incidental, as opposed to intentional, derivation and learning of new word meanings by subjects reading under circumstances that are familiar to them' (Swanborn & de Glopper, 1999, p. 262). This form of learning is particularly relevant to the investigation of story reading, an activity children likely engage in without the explicit intention to learn new words. Directed learning can take many forms, but generally it involves activities that are designed to target the acquisition of word knowledge. The relative dominance of incidental learning is illustrated by the estimation that children acquire on average 3,000 new words per year, of which vocabulary instruction in school only targets several hundred at most (Nagy, Anderson, & Herman, 1987; Nagy et al., 1985). The distinction between incidental and directed learning has recently been extended to orthographic learning in adults (Joseph et al., 2014). However, direct comparisons with directed learning are still required, as well as research with children.

Number of exposures. Effects of number or frequency of exposure have been reported for children learning to read English (Bowey & Muller, 2005; Nation et al., 2007), with better orthographic learning after a greater number of exposures. For children learning to read Hebrew, orthographic learning did not differ after one, two, or four exposures (Share, 2004). Similarly, de Jong and Share (2007) did not find an effect of number of exposures in Dutch children. This discrepancy could be due to difference in the depth of these orthographies, as Hebrew and Dutch are both considerably more shallow than English. Effects of exposure likely also depend crucially on how learning is measured, in particular whether the task can be performed on the basis of episodic memories, or requires consolidation. This issue is discussed in more detail later in this chapter.

Learning words with meaning vs. no meaning. A role for meaningful context in orthographic learning was suggested by Share (1995), who argued that contextual information may help children arrive at the correct pronunciation of words, in particular in the case of incomplete decoding attempts. However, in Share's original experiments (e.g. Share, 1999, 2004), this proposition was not tested, as novel words were always encountered in context. The role of context and semantics in orthographic learning has since been investigated with respect to several different questions. For instance, Wang et al. (2013) demonstrated that knowing a word's meaning prior to encountering it in print has a benefit for orthographic learning of word with irregular spelling-sound patterns. In line with Share's (1995) suggestion, complete decoding of irregular words is not possible, and thus semantic information could aid this process.

For words with known meanings, it has also been investigated whether encountering their written form for the first time in context affects orthographic learning compared to encountering them in isolation. Here, findings are mixed (e.g. Landi, Perfetti, Bolger, Dunlap, & Foorman, 2006; Martin-Chang, Levy, & O'Neil, 2007). In the case of words with known meanings, a beneficial effect of context has most convincingly been demonstrated for irregular words (Wang, Castles, Nickels, & Nation, 2011), where partial decoding attempts may again be guided by contextual information. Words that children do not know in spoken language when encountering them in writing constitute a fairly different case, and it is possible that semantics play a different role here. Again, mixed findings have been reported regarding potential benefits of providing semantic information alongside novel written words for orthographic learning (Nation et al., 2007; Ouellette & Fraser, 2009). Further research is clearly necessary to elucidate this issue.

Repeated vs. diverse contexts. Another variable that has received some interest in the visual word recognition literature is contextual diversity. While word frequency reflects the number of occurrences of a word, it does not take into account the nature of the surrounding context of each of these occurrences. Various measures of contextual diversity have been developed to try to capture the dissimilarity or diversity of a word's contexts (see Chapter 5). These measures have been shown to predict skilled visual word recognition on tasks such as lexical decision in adults (Hoffman, Lambon Ralph, & Rogers, 2013; Jones, Johns, & Recchia, 2012; McDonald & Shillcock, 2001), in most cases over and above the effects of word frequency. Words with higher contextual diversity show a processing advantage in lexical decision over words with lower contextual diversity. One study has also reported this effect in children (Perea, Soares, & Comesaña, 2013).

The effect of contextual diversity on word learning is less clear. For adults, beneficial effects of contextual diversity have been reported, with words presented in varying context being learned better than words presented in repeated or less variable contexts (Johns, Dye, & Jones, 2015; Jones et al., 2012). However, these observations did not always replicate (Bolger, Balass, Landen, & Perfetti, 2008). The effect of contextual diversity children's orthographic learning via story reading has not previously been examined. Here, an obvious question is whether encountering a novel word across different stories results in better or worse learning than encountering it in the same story repeatedly. This question has been investigated for pre-school (Horst, Parsons, & Bryan, 2011) and early primary school children (Wilkinson & Houston-Price, 2012), but only with respect to learning new words from spoken language.

1.4 Methodological Considerations

1.4.1 Measures of orthographic learning

To elucidate the mechanisms of orthographic learning, sensitive measures are required. The choice of these measures should be theoretically driven. From the spoken language literature, it is now clear that word learning evolves over time and can follow different time courses. Leach and Samuel (2007) proposed a useful distinction between *lexical configuration* to describe the initial acquisition of word knowledge via episode memories, and *lexical engagement* to describe the comparatively slower emergence of interactive processes between newly learned words and existing known words. This distinction maps well onto the Complementary Learning Systems account (McClelland, McNaughton, & O'Reilly, 1995). Initial exposure to novel words is said to result in episodic representations in the hippocampal memory system, underpinning lexical configuration. Through consolidation, these representations can be integrated into the non-episodic, generalised neocortical memory network, allowing the emergence of lexical engagement.

Research into the role of memory consolidation has shown that sleep can be crucial to consolidate children's initial learning of new spoken words and for markers of true lexicalisation to emerge (e.g. Henderson, Weighall, Brown, & Gaskell, 2012). While this line of research originated in the spoken language literature, it also provides a useful framework for written language learning. When investigating orthographic learning, it suggests that it is important to consider the evolving time course of word learning at an item level, which may not be possible using just a single measure. This issue is discussed in more detail in Chapter 3. Various measures of orthographic learning and their limitations are discussed next, followed by the description of a novel measure implemented and developed as part of this thesis. For the present research, the aim was to identify suitable measures of orthographic learning both in terms of lexical configuration and lexical engagement.

Orthographic choice and orthographic decision. In the self-teaching tradition, orthographic learning is commonly measured using off-line measures such as orthographic choice (e.g. Share, 1999). Children typically have to pick the correct orthographic form of a newly learned word from four choices, which include a homophone distractor and two further one-letter-different foils. The homophone distractor ensures that phonological decoding alone is insufficient to succeed at this task, so children have to use additional knowledge of the correct orthographic form. This task is quite removed from natural reading, as children are simultaneously presented with two homophone words, which might be confusing (Castles & Nation, 2008). Furthermore, it infers orthographic knowledge indirectly, rather than directly tapping children's ability to automatically and rapidly recognise words. Thus it might be subject to strategic influences. If children have no representation of the correct orthographic form, they have a 25% chance of picking the correct answer. This is further increased to 50% if they are able to correctly identify the two possible phonological forms (Wang et al., 2011). This could allow some children to strategically rule out two of the four answers. This last criticism has been addressed by a more recent variant of the task, the orthographic decision task (Wang et al., 2011). It uses the same stimuli as the orthographic choice task, but they are presented individually in random order, requiring a yes/no forced-choice decision. While this might be preferable to orthographic choice, it still only measures lexical configuration, rather than the extent to which words have been lexicalised. It also requires children to inhibit responding affirmatively to the homophones, which have the correct phonological form, which some children could find more difficult than others.

Spelling. Spelling is another measure of lexical configuration commonly used which explicitly taps children's knowledge of word forms. Unlike orthographic choice and orthographic decision, the task requires recall rather than recognition. Orthographic decision and spelling have been found to be correlated (e.g. r =. 61 for regular words, Wang et al., 2013). However, performance on spelling and orthographic choice has also been found to differ within the same experiment (Kyte & Johnson, 2006; Wang et al., 2011), suggesting that the two measures are not entirely redundant.

Naming time and fixation duration. Naming time offers an indication of the speed with which children recognise words and is thus arguably closer to the process of interest than the above measures. Following exposure to novel words, naming times are compared for the target word spelling vs. a homophone variant, and orthographic learning is inferred if the correct target word spellings are read faster than the homophones (e.g. de Jong & Share, 2007; Share, 1999). A novel measure related to naming time are fixation durations recorded during reading, which allows word reading times to be assessed without the requirement of spoken output. In general, it is assumed that the location where a reader is looking reflects what is currently being processed, and that greater processing demands require longer fixation times when words are read (Rayner, 1998). Orthographic learning can be inferred if fixation times become shorter as novel word forms become familiar (Joseph et al., 2014). By tapping reading processes online as they unfold, this measure is much closer to word recognition as it happens during natural reading than the offline measures discussed above. A range of different eye-movement measures, i.e. the duration of the initial fixation of the word as well as re-reading times, allows an insight into the word recognition process at different stages that cannot be accessed by offline measures. However, it is difficult to directly infer lexical engagement of novel words from reduction in fixation times.

Measuring naming time or fixation duration also allows for investigation of length effects. Length effects, e.g. increasing naming times with increasing word length, have been observed in adults for nonwords and unfamiliar or low frequency words, but less so for familiar or high frequency words (reviewed in Kwok & Ellis, 2015). Applied to orthographic learning, this suggests that as written forms become familiar, a reduction in length effects should be observed. This has indeed been reported in adults following repeated exposure to nonwords (e.g. Kwok & Ellis, 2015). While length effects have been observed in children (e.g. van den Boer, de Jong, & Haentjens-van Meeteren, 2013), they have yet to be applied to children's orthographic learning of novel words. This direction was not chosen for the present research, as it does not allow for the assessment of lexical engagement.

Lexical competition. While lexical configuration is typically assessed by measuring explicit knowledge of word forms, lexical engagement measures have to tap emerging interactions between a novel word and known words. In turn, such interactions are considered evidence that a novel word has been lexicalised. A specific type of interaction, lexical competition, has been proposed as a sensitive indicator of lexical engagement (Qiao & Forster, 2013; Qiao, Forster, & Witzel, 2009).

Bowers, Davis, and Hanley (2005) pioneered the use of lexical competition following exposure to novel written words, although not in the context of orthographic learning. Adults were trained on nonwords (e.g. *banara*), which were one-letter-different from familiar 'hermit' words (e.g. *banana*), i.e. words without existing neighbours. During training, participants repeatedly typed novel words (e.g. *banara*). Performance on a semantic categorisation task (artefact vs. natural object) was compared for hermit words (e.g. *banana*) to which new neighbours had been introduced (e.g. *banara*) vs. hermit words to which no new neighbours had been introduced. Immediately after training, participants showed a weak interference effect on hermit words with newly learned neighbours, e.g. adults took longer to categories *banana* as a natural object following exposure to *banara*. This interference effect increased after a 24-hour delay. From this it was inferred that participants had lexicalised the new words, which then engaged in lexical competition with their neighbours.

However, Qiao et al. (2009) criticised this study by arguing that the interference may have been due to an episodic trace left by the repeated typing of *banara*, rather than lexical competition (for further discussion of this issue, see Chapter 6). In other words, they questioned whether the novel words were indeed lexicalised. They proposed the prime-lexicality effect as a better behavioural marker of lexicaltion than interference on semantic categorisation. This effect, observed on masked priming lexical decision, is argued to clearly differentiate words from nonwords. The general format of this task is the standard three-field paradigm of Forster and Davis (1984). A forward mask consisting of a row of hashtags (#) is followed by a briefly presented prime (around 50 ms) in lower case letters. This is followed by the target in upper case letters, to which a lexical decision response is made. Whereas nonword neighbours facilitate responses when used as primes (e.g. anple – APPLE), real word neighbours do not facilitate or even inhibit responses (e.g. ample – APPLE; Davis & Lupker, 2006). In this respect, orthographic neighbours are defined to differ in one letter only that can be substituted (Coltheart, Davelaar, Jonasson, & Besner, 1977). The general explanation for the prime-lexicality effect is that when reading a word, its orthographic neighbours are also partially activated. On a primed lexical decision task, word primes that are neighbours compete with the target words, leading to slower responses compared to unrelated primes (see Davis & Lupker, 2006; Qiao & Forster, 2013). In contrast, faster responses to target words primed by one-letter-different nonwords is explained by additional activation of from provided by the orthographically similar nonword primes; by definition, nonwords are not in the lexicon and thus cannot compete.

Qiao et al. (2009) and Qiao and Forster (2013) used this effect with some success to measure novel word lexicalisation in adults. Participants completed the same typing task as used by Bowers et al. (2005). However, instead of the semantic categorisation task, after training they completed a masked priming lexical decision task, as described above. In this case, the lexicality comparison is made between trained vs. untrained nonwords (analogous to words vs. nonwords) serving as primes. As in the paradigm described above, lexical decisions are made on target words (e.g. *banana*) that are primed by one-letter-different nonwords (e.g. *banara*). If nonwords become lexicalised via training, there should be a reduction in the priming for trained nonword primes (due to lexical competition) compared to untrained nonword primes. Qiao et al. (2009) found no evidence for this and concluded that the repetitive typing task was insufficient for lexicalisation of the nonwords. In particular, this training provided no information about the meaning of the words. In a follow-up study, training was designed to explicitly promote word learning and introduce meanings for the new words (Qiao & Forster, 2013). In Experiment 1, participants completed a single learning session, which involved exposure to the new words alongside pictures and definitions, word-picture matching practice, picture-word matching, and word-definition matching. Afterwards, priming effects on lexical decision were once again compared for trained vs. untrained nonword primes. As in Qiao et al. (2009), facilitatory priming was found for both, indicating that despite the more elaborate training paradigm, trained nonwords had not been lexicalised and competition not induced. In Experiment 2, training was extended to four sessions following the same procedure as in Experiment 1. This time, trained nonword primes no longer facilitated responses to their target neighbours, showing that lexical competition could indeed be induced. It remains unclear whether the inclusion of semantics was essential to fully establish lexicalisation, as semantic training was not compared with non-semantic training. It might be that the emergence of lexicalisation was solely driven by the extended training period, e.g. due to the greater number of repetitions of the orthographic forms, or benefits of multiple, spaced sessions in Experiment 2 compared with Experiment 1.

Similar to Bowers et al. (2005), Bakker, Takashima, van Hell, Janzen, & McQueen (2014, Experiment 3) also demonstrated the inhibitory effects on a semantic categorisation task (‘Is this an artefact or a natural object?’). Participants were exposed to novel words via a letter monitoring task, where they had to indicate the presence or absence of specific letters. After 24 hours, interference by the trained novel words on semantic categorisation was observed. This effect co-occurred with a decrease in recognition memory of the trained nonwords, indicating only weak episodic memory traces, thus arguing against the episodic account proposed by Qiao et al. (2009). It is noteworthy that contrary to Qiao et al. (2009), competition effects emerged in the absence of semantic training. Because of this, Bakker et al. (2014) claimed that masked priming ‘may simply be less sensitive to the small effects of early lexicalisation than tasks like pause detection or semantic decision’ (p. 127). They also reported a failure to find significant competition effects when a non-semantic progressive demasking task was used during pilot testing. Unfortunately, this was not described further. It is clear that research is needed to clarify the sensitivity of the prime-lexicality effect vs. other measures of lexicalisation and lexical competition.

While there is promising research using the prime-lexicality effect as a measure of learning in adults, it has yet to be applied to developmental research in orthographic learning. However, there are reasons to be optimistic about its utility. Like adults, children in Grade 5 show facilitatory priming by one-letter-different nonwords for word with low neighbourhood density, i.e. few orthographic neighbours (Castles et al., 2007). Given that hermit words such as *banana* constitute a case of extremely low neighbourhood density, it would be expected that like adults, children show facilitatory form priming for such words. This can be used a baseline comparison for lexical competition effects induced via training. As in adults, for novel words that become lexicalised, a reduction in facilitation or the emergence of inhibition would be expected.

A clear advantage of lexical competition measures is that RT is a continuous measure. Orthographic learning of novel words, or the change from nonword to word status, is probably also a continuous process, as discussed before with regard to consolidation of newly learned spoken words in memory (e.g. Gaskell & Dumay, 2003). Neither lexical configuration nor engagement of novel words is likely to be fully captured by binary measures reflecting only the presence vs. absence of learning (e.g. accuracy of 0 vs. 1 on orthographic choice). Instead, there are likely gradual differences both in the accumulation of knowledge about new words and their integration in the lexicon (see Frishkoff, Collins-Thompson, Perfetti, & Callan, 2008, for further discussion). Words have been argued to show gradual variation in lexical quality, i.e. in the quality of their orthographic, phonological, and semantic constituents (Perfetti & Hart, 2002). Indeed, it is well-documented that skilled readers do not process all written words with equal efficiency. For example, adult performance on lexical decision is affected by factors such as word frequency, imageability and age of acquisition, to name but a few variables (for a review, see Balota, Cortese, Sergent-Marshall, Spieler, & Yap, 2004). Differences in skilled visual word recognition are determined using continuous measures such as RT, reflecting gradual variation in processing. In contrast, relatively little is known about which factors influence the development of skilled word recognition. To fully elucidate this issue, experiments that are sensitive to gradation in learning and lexicalisation are required.

1.5 Thesis Aims and Overview

The research reported in this thesis aimed to examine learning under conditions comparable to children's everyday reading experience. Thus, exposure conditions should be similar to the opportunities children have for orthographic learning in their day-to-day life. This concerned both the reading materials and as well as the items to be learned. Much of the research in this thesis focussed on incidental learning via story reading, an activity children regularly engage in (Clark, 2012). Stories were written to be motivating, involving recurrent characters similar in age to the children taking part. With regard to items, existing research often uses artificial nonwords as a proxy for unfamiliar words, including when researching children. This is ethically problematic and can be avoided given the large number of words unknown to children. Throughout the present research, children were exposed to unfamiliar low-frequency words rather than nonwords.

To investigate learning within this framework, a first aim was to develop a sensitive measure of orthographic learning. Based on the methodological considerations above, multiple measures of orthographic learning were contrasted in the experiments, to allow for comparisons. Orthographic decision and spelling were chosen to measure lexical configuration, i.e. how well children learned the form of the novel words. In all experiments, semantic learning of novel words was also assessed, using definition knowledge. Given that lexical engagement of novel written words has not been studied in children, the first experiment employed two novel measures of lexical engagement: the prime-lexicality effect in lexical decision and interference effects in semantic relatedness judgements. Based on the results of this experiment, only the prime-lexicality effect was used in subsequent experiments.

A second aim was to investigate factors influencing orthographic learning in typically developing children, addressing a range of theoretical questions. Each experiment contrasted two exposure conditions, which manipulated a specific factor hypothesised to influence learning outcomes. Chapter 2 reports an experiment investigating the role of incidental vs. directed exposure. Chapter 3 reports an experiment on the effects of number of exposures. In Chapter 4, two experiments investigated the role of semantics, in particular the presence vs. absence of context. The experiment reported in Chapter 5 further investigated the nature of the story contexts by manipulating their diversity. Together, this research elucidates children's learning of novel words via reading experience.

Chapter 2:

The Prime-Lexicality Effect as a Novel Measure of Children's Orthographic Learning

Experiment 1

2.1 Introduction

Chapter 1 outlined the need for sensitive measures of orthographic learning. Addressing this need will enable research to better elucidate underlying mechanisms by investigating the impact of factors of interest on learning outcomes (see Nation & Castles, in press). The prime-lexicality effect (e.g. Qiao & Forster, 2013) was identified as having potential, but has yet to be used to measure the emergence of lexical competition in children's reading development. Experiment 1, reported in this chapter, asked whether the prime-lexicality effect could capture word learning in a training paradigm with children. The utility of a second task as a measure of learning-induced competition was also assessed. Orthographic learning of novel words was assessed comprehensively, by measuring children's knowledge of novel words in the tradition of the self-teaching hypothesis, in addition to the measures of lexical competition. Two exposure conditions were compared to investigate their effect on word learning and lexicalisation, thus allowing a better understanding of the factors involved in orthographic learning.

2.1.1 Investigating the lexicalisation of novel written words

Traditionally, children's orthographic learning has been measured largely by directly assessing their knowledge of novel word forms by asking them to recognise or recall the correct spellings. As outlined in Chapter 1, this approach has led to some important theoretical discoveries, but has its limitations. The prime-lexicality effect offers a promising new approach to measuring orthographic learning by tapping the lexicalisation of novel written words. Experiment 1 constituted a first attempt at demonstrating this effect in a training study of children's orthographic learning. Only two learning studies using a similar paradigm have been published (Qiao et al. 2009, Qiao & Forster, 2013), both with adults, so little is known about the conditions under which novel written words become sufficiently lexicalised to exhibit 'word-like' priming effects. Lexical competition effects of newly learned written words have not been investigated following exposure via reading new words embedded in text in children or adults. The present experiment assessed whether lexical competition between newly-acquired words and their known neighbours would emerge in children following a training paradigm based on reading meaningful text, or whether tasks focussed on directed word learning are necessary.

Qiao and colleagues (Qiao et al., 2009, Qiao & Forster, 2013) conducted three experiments investigating the emergence of a prime-lexicality effect following exposure to nonwords, discussed in Chapter 1. Qiao et al. (2009) did not find lexical competition following a single training session where participants repeatedly typed a set of nonwords. Instead, priming effects of trained nonwords remained comparable to untrained nonword primes – i.e. facilitatory. The authors suspected that a deeper level of processing than that evoked by simple typing might be necessary for lexicalisation. Consequently, in Qiao and Forster (2013), participants were exposed to word forms, corresponding pictures and definitions, and performed word-picture matching and word-definition matching tasks. After a single training session, no prime-lexicality effect was found (Qiao & Forster, 2013, Experiment 1). However, when the same training regime was extended to four sessions spread over two weeks, a prime-lexicality effect did emerge (Qiao & Forster, Experiment 2). It remains unclear precisely which aspect of training was critical for successful lexicalisation - the enhanced depth of processing in combination with multiple sessions, or just the latter. For instance, it was not investigated whether multiple sessions of a typing task without the provision of meaning would yield a similar result.

All of the above training paradigms are very different from children learning about novel words when coming across them in their reading activities. In Qiao and Forster's (2013) experiments, participants were instructed to pretend that they were learning novel English words, making them aware that they were dealing with nonwords. During training, participants were also intermittently tested on their emerging semantic knowledge, rendering word learning an explicit goal of training. The present study compared a training paradigm based solely on reading novel words in stories with a paradigm involving additional directed engagement with the novel words, the latter being the one more similar to Qiao and Forster (2013). Instead of using artificial nonwords, children were exposed to rare English words.

2.1.2 Incidental orthographic learning

As discussed in Chapter 1, exposure to meaningful text provides a powerful learning opportunity. Children have been shown to acquire both word meanings (e.g. Nagy et al., 1985; Swanborn & DeGlopper, 1999) and word forms (e.g. Share, 1995) when encountering novel words in their reading experience. Learning about words from text and without directed instruction has been termed incidental learning (e.g. Ender, 2014). For the present purpose, incidental learning is defined within the context of reading, where the intention of word learning is not made explicit and the only opportunity to learn comes from encountering novel words in text. My focus is orthographic learning, and here, existing research on incidental exposure is relatively sparse both with regard to theory and experimental evidence, which will be discussed in turn.

Regarding underlying theory, the self-teaching hypothesis suggests a mechanism for incidental learning of word forms: spelling-to-sound decoding. However, decoding success does not fully account for the variance in orthographic learning (e.g. Nation et al., 2007). This suggests that factors beyond phonological decoding influence learning outcomes, but what these factors might be requires further investigation (see Nation & Castles, in press, for further discussion). With regard to incidental learning, the amount of exposure to novel words could influence learning. This will be discussed in detail in Chapter 3. The availability and nature of context might also be important, but this aspect of the self-teaching hypothesis is underspecified, an issue discussed in greater detail in Chapters 4 and 5. Further experimental research is clearly necessary to formulate comprehensive mechanisms of incidental orthographic learning.

Specifying a mechanism for directed learning of novel word forms is also a challenge, as there is a seemingly infinite number of ways in which children could be trained on novel written words. In order to devise a paradigm close to children’s real-life experiences, the present study focussed on directed learning as it could happen in the classroom. Here, it seems likely that teaching activities focussed directly on the novel written words would aim to boost learning of their spellings and/or meanings.

Any activity that draws attention to the precise spelling of a novel word, i. e. its individual letters in order, could potentially support orthographic learning. This argument was previously made by Landi et al. (2006), who compared 7-year-old’s orthographic learning of words read in isolation vs. in sentence contexts. Target words were high in frequency and highly predictable from their sentence contexts. Orthographic learning was better for words read in isolation than in context. The authors argued that this was likely due to increased attention to word spellings in the isolation condition, as children had to read the words without the help of constraining sentence contexts. Arguably, a more direct way to increase processing of individual letters is to ask children to copy the spellings of novel words. This procedure has previously been used with adults (Bowers et al., 2005).

Directed exposure aimed to increase semantic learning of novel words could involve activities such as asking children to generate the meanings of novel words following exposure. Generative activities following exposure to text have been shown to boost semantic learning of novel words in adult second language learners (Joe, 1998). In general, one could argue that better semantic knowledge contributes to higher overall lexical quality of newly acquired words. There could also be a knock-on effect of emerging semantic knowledge on orthographic learning. This issue is discussed in greater detail in Chapter 4. An example of directed exposure to word meanings in an orthographic learning study with children is the use of pictures and oral definitions (e.g. Ouellette & Fraser, 2009; Wang et al., 2011).

Taken together, the types of directed exposure described above could be argued to result in better orthographic and semantic representations of newly learned words relative to incidental exposure. Directed exposure could increase not only the chance of acquisition in the first place, but also the lexical quality of the resulting entry in the lexicon. In other words, even if both incidental and directed exposure resulted in equal learning rates on a binary measure (e.g. accuracy scored as 0 or 1), lexical quality might nonetheless differ. Thus it is important to investigate which measures are best able to capture gradual differences in lexicalisation, and by extension, lexical quality. It is yet unknown to what extent lexical competition as measured by the prime-lexicality effect is sensitive to subtle, continuous differences in orthographic learning and lexical quality, an issue explored throughout this thesis.

Orthographic learning following incidental exposure to novel words in text has been demonstrated in children within the context of the self-teaching paradigm (e.g. de Jong & Share, 2007; Nation et al., 2007; Share, 1999), and in studies of written word learning in adults (Bolger et al., 2008; Joseph et al., 2014). However, much of this research has relied heavily on the use of explicit recall or recognition tasks to measure learning, e.g. orthographic choice or spelling. These measures indicate the extent to which a child has acquired explicit knowledge of novel word forms ('lexical configuration'). It remains unknown whether incidental exposure to novel written words can give rise to lexicalisation as measured via lexical competition effects on neighbours in the lexicon ('lexical engagement', see Leach & Samuel, 2007, Chapters 1 and 3 of this thesis). An alternative measure of learning was implemented by Joseph et al. (2014), who used changes in fixation duration to indicate orthographic learning in adults following silent reading of novel words embedded in sentences (see Chapter 1). Eye-movements offer a window into the on-line processing of written words, but like orthographic choice, they do not directly tap lexicalisation. A further methodological concern regarding the adult literature is the frequent use of single-sentence contexts (Bolger et al, 2008; Joseph et al., 2014). It is questionable whether this type of exposure can be termed truly incidental, as unfamiliar words appeared in every single sentence, potentially making them unnaturally salient. Furthermore, most previous studies have used artificial stimuli and nonwords (e.g. *banara* introduced as a neighbour of *banana*, Qiao & Forster, 2013), which may not be comparable to the processing of words encountered in natural reading outside the laboratory. The present experiment addressed these issues by using multi-sentence story contexts to approximate incidental learning of unfamiliar words instead of nonwords. Learning was assessed in terms of both explicit orthographic knowledge of novel words, and their lexicalisation via lexical competition effects.

One way to investigate the efficacy of incidental learning is to compare it to learning following directed, intentional exposure. Directed exposure was used in Qiao and colleagues' experiments using lexical competition (Qiao et al., 2009; Qiao & Forster, 2013), and in some studies assessing explicit orthographic learning (e.g. Ouellette & Fraser, 2009), but has rarely been contrasted with incidental exposure in the same experiment. Bolger et al. (2008, Experiment 1) compared novel word learning in adults following sentence reading in the presence or absence of definitions. Although not explicitly framed as such, this could be considered a comparison of directed vs. incidental learning. There was no effect of definitions on orthographic choice accuracy or RT. However, open questions remain as it is unknown whether this would also apply to developing readers, or to a different implementation of an 'incidental vs. directed exposure' manipulation. The present experiment compared incidental learning with a more directed approach in children.

In summary, so far it has been shown that some orthographic learning is possible following incidental exposure, but measures have not yet been extended to lexical competition. It remains unknown if incidental exposure and directed exposure result in equal learning in children, and if results differ depending on how learning is measured. The present experiment aimed to investigate these questions comprehensively by using novel measures of lexical competition in addition to measures of orthographic learning from the self-teaching literature. Addressing these issues has important implications for educational practice, as it could clarify whether children who are already able to decode benefit from directed exposure to novel word spellings, beyond encountering them in their independent reading activities.

2.1.3 The present experiment

Experiment 1 compared two different learning conditions to contrast incidental and directed exposure. Children in the reading-only condition read aloud stories containing novel words over three exposure sessions. Children in the reading-plus condition read aloud the same texts as in the reading-only condition, with two additional directed elements to promote learning of both word forms and meanings – spelling and meaning generation. The total number of exposures to each novel word form was constant across conditions.

Learning of the novel words was assessed in both in terms of explicit knowledge of their form and meaning, and with regard to their lexicalisation. Explicit knowledge of form was assessed using orthographic decision (an adaptation of orthographic choice, see Chapter 1 for further discussion) and spelling. Explicit knowledge of meanings was assessed via definition knowledge. Lexicalisation of novel words was assessed via the emergence of lexical competition. The first measure for this was the prime-lexicality effect, which was measured by comparing performance on a masked priming lexical decision before and after exposure to the novel words. Prior to exposure, novel words were expected to exhibit nonword-like facilitatory priming effects. After exposure, the absence of facilitation or the emergence of inhibition would indicate lexical competition between newly lexicalised words and their neighbours.

Given the debate in the literature regarding which task is most sensitive to lexical competition effects (see Chapter 1), in addition to the prime-lexicality effect, a second measure was included, which assessed lexical competition on semantic judgments rather than lexical decision. The constraints of using words rather than nonwords made it too difficult to design a semantic categorisation task analogous to Bowers et al. (2005) and Bakker et al. (2014, see Chapter 1), a point reflected on in more detail in the General Discussion. Hence a novel task was created based on a semantic relatedness judgements task, developed by Rodd et al., (2012). This task has not previously been used to assess lexical competition. Like masked priming lexical decision, the semantic relatedness task was administered before and after exposure to the novel words to measure the emergence of competition effects against a baseline.

Using these two measures, Experiment 1 aimed to establish lexical competition as a novel, sensitive measure of orthographic learning in children. Given that the reading-plus condition involved directed focus on the spelling and meaning of the novel words, it was anticipated that learning of form and meaning would be superior in this condition.

2.2 Method

2.2.1 Design

The experiment involved a mixed design with a between-subjects manipulation of exposure condition (reading-only vs. reading-plus). The experiment consisted of three phases – pre-exposure baseline, exposure phase, and post-exposure test. Data for masked priming lexical decision and semantic relatedness tasks were collected before and after, yielding a further within-subject factor of test session (pre-exposure vs. post-exposure). Spelling, orthographic decision, and definition knowledge data were collected once after exposure.

2.2.2 Participants

Thirty-one children (16 female, 15 male) from Year 5 and Year 6 classes participated in the experiment. They were aged between 9.25 and 11.42 years (*M* = 10.13, *SD* = .72). They were selected on the basis of being monolingual native English speakers scoring no more than 1*SD* below the standardised average on the Test of Word Reading Efficiency (TOWRE; Torgesen, Wagner, & Rashotte, 1999). This test consists of two subtests – word reading and nonword reading. Performance on the two subtests was highly correlated (*r* = .85). One child was excluded based on poor reading ability. Data from a further child were excluded because the child did not complete all sessions. The final sample consisted of 15 children in the reading-only group and 14 children in the reading-plus group. Descriptive statistics are shown in Table 2.1. Raw TOWRE scores indicate reading ability independent of age, which was used in subsequent analyses. Standard scores reflect reading ability standardised for age, which are reported here to illustrate the nature of sample relative to the general population. The overall mean total standard score was 111.72 (*SD* = 13.68), indicating that children performed towards the upper end of the population average. Groups did not differ significantly in age or reading ability (see Table 2.1). Children were recruited using an opt-out procedure, where parents were sent two information sheets over the course of two weeks with the opportunity to opt out of their child’s participation in the study by returning a response slip. Ethical approval was obtained for all experiments reported in this thesis.

Table 2.1. Descriptive characteristics of reading-only and reading-plus groups as well as group comparisons (independent-samples t-tests)

	Reading-only (*N* = 15)		**Reading-plus (*N* = 14)**		**Group comparison**
Measures	*M*	*SD*	*M*	*SD*	*p*
Age (years)	10.12	0.71	10.14	0.75	.940
TOWRE total raw score[a]	114.87	13.54	108.5	14.09	.225
TOWRE total standard score[b]	114.07	13.08	109.21	14.35	.349

Notes. [a]total number of words read correctly on both subtests in 45 seconds per test, [b]*M* = 100, *SD* = 15.

2.2.3 Materials and procedure

2.2.3.1 Novel words

Twenty word pairs comprising a novel word and an orthographic neighbour base word were selected. Each pair met the following criteria:

a) neither item has further orthographic neighbours likely to be known by children, differing by a single letter substitution in the CELEX database (0-1 additional neighbours, *M* = 0.3 (CELEX frequency *Min* = 0.63, *Max* = 16.15, *M* = 4.08), e.g. *yield*
b) each novel word was relatively low in frequency (CELEX frequency *Min* = 0, *Max* = 2.29, *M* = 1.17), e.g. *wield*

c) each base word was relatively high in frequency (CELEX frequency *Min* = 1.34, *Max* = 2496.31, *M* = 196.11), e.g. *field*

Ten novel words, ten base words and two additional neighbours were piloted on a sample of 16 children (mean age 7.88 years, *SD* = 1.41) who took part in a half term science week. Children had English as their dominant language. Standardised average word reading ability as assessed by the TOWRE was at the top end of the population average (*M* =115, *SD* = 41.61). Children were shown each word on a flashcard. They were asked to read the words aloud, indicate whether they had heard them before, and provide a definition. Summary pilot data for each word type is shown in Table 2.2.

Table 2.2. Mean (*SD*) performance on novel words, base words, and additional neighbours in the pilot test.

	Novel words	**Base words**	**Additional neighbours**
Reading accuracy	.70 (.46)	.97 (.17)	.66 (.48)
Yes response to "Have you heard this word before?"	.14 (.35)	.97 (.17)	.07 (.26)
Definition knowledge	.04 (.19)	.96 (.19)	0 (0)

Pilot data confirmed that children had very good knowledge of the meanings of the base words and very minimal knowledge of the novel words. Additional neighbours not included in the chosen word pairings were also unknown by the children. Reading accuracy was lower for novel words than base words, which is in line with literature showing children read words more accurately if they know their meaning than if not, and if they are of higher frequency than lower frequency (e.g. Nation & Snowling, 1998)

For counterbalancing purposes, the 20 pairs of novel words and base words were divided into two lists of 10 pairs. Children were allocated to either list based on the order in which they participated in the experiment. All of the word pairs are listed in Appendix 2.1.

2.2.3.2 Pre-exposure baseline assessments

Due to scheduling constraints in school, some children completed these tasks on the same day as the start of the exposure phase, whereas others completed them on a separate day. All computerised tasks in the experiment were presented on a Dell Latitude E6400 laptop controlled by E-Prime 2.0 software (Schneider, Eschman, & Zuccolotto, 2012).

Masked priming lexical decision. This task tapped the prime-lexicality effect. As shown in Figure 2.1, each trial comprised a fixation cross shown for 500 ms followed by a forward mask (########) displayed for 500 ms. The prime was shown in lowercase letters for 57 ms, followed by the target in uppercase letters. The target was presented until children responded with a button press. Children were asked to indicate whether words were ‘real’ or ‘made-up’. ‘Real word’ responses were indicated by pressing the ‘m’ key on the keyboard, ‘made-up word’ responses by pressing the ‘z’ key.

Children made lexical decisions to 40 targets: 20 words and 20 nonword foils. Each target was primed by a masked word. For each child, 10 of the words were the base words from their allocated list; the other 10 words were base words from the counterbalanced list. The 10 base words from the child’s allocated list were primed by their novel word neighbours, which the child subsequently experienced in the exposure phase. The 10 base words from the counterbalanced list were primed by 10 unrelated[1] low frequency control words, all of which also had no more than one orthographic neighbour. The 20 nonword foils each differed from an English word by one letter. Half of the nonwords were primed by related, low-frequency word primes (orthographic neighbours), and half were primed by unrelated, low-frequency word primes. The full list of prime-target pairs can be found in Appendix 2.1.

In the first session, children completed a practice version with eight trials, where feedback was given. The experiment itself started with four dummy trials at the start before the 40 items were presented in a random order.

[1] Unrelated was defined as more-than-one-letter different, i.e. non-neighbouring

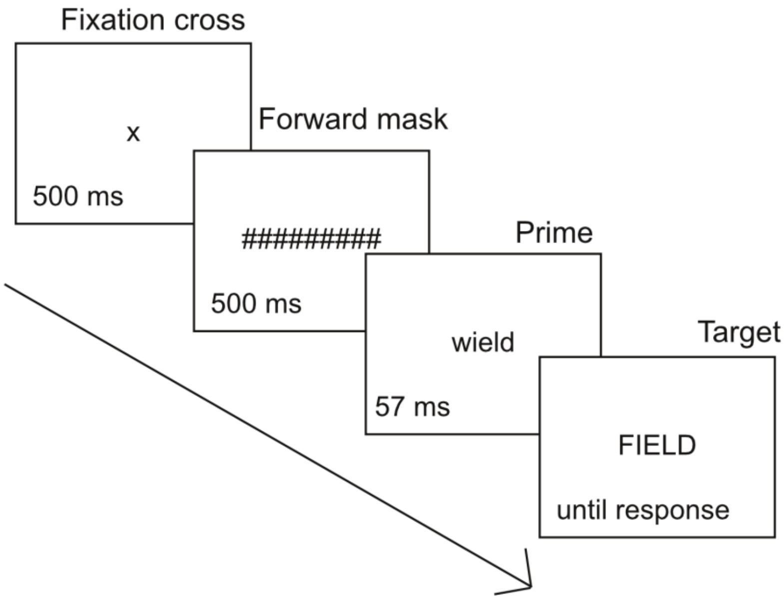


Figure 2.1. Example of a related word trial in masked priming lexical decision.

Semantic relatedness. To assess competition effects on semantic relatedness judgements, a task was developed based on Rodd et al. (2012). Children read aloud a sentence from the screen (e.g. ‘The pitch behind the school was used for football games.’). Immediately following this, the experimenter manually initiated the appearance of a probe word below the sentence (e.g. ‘field’). Children were asked to press a button to indicate as quickly as possible whether or not the probe word was related to the sentence meaning. For the 20 YES-trials, the probe word constituted the 20 base words from both lists. Hence, for each child, 10 were the neighbours of novel words, and 10 were neighbours of control words. For the 20 NO-trials, the probe words were fillers of similar length and frequency to the base words. Ratings from 49 adult native English speakers were collected to ensure the classification of sentence-probe pairings as related/unrelated was appropriate. Sentences and probes can be found in Appendix 2.2.

YES responses were indicated by pressing the ‘m’ key on the keyboard; NO responses were indicated by pressing the ‘z’ key. RTs were measured from the onset of the probe word until a response was made. Pre-exposure baseline measurements were compared against post-exposure measurements on the same task. Longer RTs to neighbours of novel words post-exposure compared to pre-exposure would indicate interference or competition effects, while practice effects were expected for control words after exposure.

2.2.3.3 Exposure phase

The exposure phase ran across three sessions completed on separate days (see Table 2.3). The number of days across which children completed the exposure phase varied between 3 and 9 days (M = 5.62, SD = 1.76). This did not differ between the two groups (p = .079).

Story contexts. Twenty story contexts were created – one per novel word. To create a coherent narrative akin to chapters in a book, all stories revolved around of two children who had to find and free words stolen by an evil professor. Each story involved a different hiding place where the protagonists had to ‘crack a code word’ by finding out about the meaning of the novel word embedded in the story. Each story referred to the meaning of the novel word in six different ways – one dictionary-style definition (e.g. ‘to wield a weapon means to hold it out ready to attack’), three sentences conveying additional semantic information (e.g. ‘a knight might draw his sword and wield it at a dragon’) and two further mentions without additional semantic cues (e.g. ‘I see you understand what wield means’). Chapters were laid out on A5 story cards modelled on the typeset of age-appropriate novels, which contained a coloured illustration of the novel word that fitted with the story. Chapters were matched on length (M = 128 words). Full stories can be found in Appendix 2.3.

Exposure conditions. Children were randomly allocated to either the reading-only or the reading-plus condition.

Reading-only condition. In session 1, children were told they would be reading a story printed on several cards similar to chapters in a book. They were given a short prologue to read which explained the premise of the stories. They were then asked to shuffle the cards face down to determine the order in which stories were read. They were asked to then read each story aloud to the experimenter. At the end, they read a short epilogue. In session 2, children were told they would be reading the same stories again, but this time they could decide the order in which they were read by choosing from the story cards. Children read the stories aloud as in session 1, but without prologue and epilogue. In session 3, children were given a map, which showed the locations used in the story cards. They were asked to decide on a route for the protagonists to take by connecting the locations on the map with a line. The story cards were then read in this order as before. Children were corrected if they read out the novel words incorrectly and were generally assisted in their reading as required.

Reading-plus condition. In session one, children read the stories as in the reading-only condition. In session two, after reading each story in the order of their choice, they were asked to write down each novel word on a card and to provide their own definition underneath. Children could refer back to the story cards. Spelling mistakes were corrected for the novel words only. In session three, after completing the map, children were given a modified version of the story cards where all the novel words were missing and needed to be filled in. They were asked if they could remember the code word from the story based on the picture and the story context. If they could not remember it, the experimenter provided its spoken form. The experimenter then wrote the novel word down on a separate sheet to remind them of the correct spelling. Children were then asked to copy down the novel words into all the gaps and then they read the completed card aloud. In sessions one and two, the experimenter provided the written form of the novel words, providing an additional exposure per novel word. For these sessions, the story cards had only five exposures instead of six per novel word to match the total number of exposures across the reading-only and reading-plus conditions (Table 2.3). Children were corrected if they read out the novel words incorrectly and were generally assisted in their reading as required.

Table 2.3. Number of exposures in text as well as in additional tasks for two exposure conditions.

	Reading-only condition	**Reading-plus condition**
Session 1	6 exposures in text	6 exposures in text
Session 2	6 exposures in text	5 exposures in text + 1 in definition task
Session 3	6 exposures in text	5 exposures in text + 1 in fill-in-the-gaps task
Total number of exposures	18	18

2.2.3.4 Post-exposure assessments

These assessments were administered in the order listed below. Children were tested between 1 and 6 days after completing the exposure phase (*M* = 1.48, *SD* = 1.06), depending on their availability. The test delay did not differ between the two groups (*p* = .795).

Word definition recall. Children were asked to verbally define each novel word in the context of a 'code breaker game', presented in E-prime 2.0. The written form of each item appeared one-at-a-time on the computer screen and responses were scored by the experimenter as correct if at least one critical feature or function was produced.

Masked priming lexical decision. Children completed the lexical decision task as before, without the practice version.

Semantic relatedness. Children completed the semantic relatedness task as before, without the practice version.

Orthographic decision. This task (based on Wang et al., 2013) assessed how well children learned the forms of the novel words. A nonword homophone distractor was created for each novel word (e.g. *wield* – *whield*) as well as two nonword distractors differing in one consonant (e.g. wielk, wielp). Care was taken to maintain length, but some homophone distractors needed to differ in length from the novel words. Items were presented individually one at a time on a laptop, using E-Prime 2.0. Children performed this task on the novel words and distractors from their allocated list, i.e. 10 novel words and 30 distractors. Children were asked to indicate whether or not the shown word was one of the code words from the story in the correct spelling. YES responses were indicated by pressing the 'm' key on the keyboard; NO responses were indicated by pressing the ‘z’ key. They were told to look carefully and take their time with each decision. For full list of stimuli, see Appendix 2.4. The task was scored based on Wang et al.’s (2013) procedure. For each novel word, performance was coded as accurate if a YES response was made to the correct spelling and all three incorrect spellings were rejected.

Spelling. The ten novel words were dictated from the appropriate counterbalanced list in a random order and children were asked to write them down. Responses were scored as correct or incorrect.

2.3 Results

Where possible throughout this thesis, linear mixed effects models with maximal model structure (see Barr, Levy, Scheepers, & Tily, 2013) were computed using the *lme4* package (version 1.1-7; Bates, Mächler, Bolker, & Walker, 2015) in *R* (version 3.1.0; R Core Team, 2014). By-subject and by-item random intercepts and random slopes were included for each fixed main effect and interaction. Fixed effects were centred around their mean to minimise collinearity. Models which failed to converge with maximal structure were simplified by removing random effects explaining the least amount of variance in the outcome variable, or by removing random correlations. For models with continuous outcome variables, significance was determined using the cut-off point of $|t| > 2$, based on recommendations by Barr et al. (2013). For models with binary outcome variables, mixed effects logistic regression models were computed. Here, the cut-off point for significance was $p < .05$. To control for individual variation in word reading ability, children’s total raw TOWRE scores were included as covariates in the analyses.

2.3.1 How well did children learn the novel words?

Three tasks assessed how well novel words were learned. Orthographic decision, and spelling assessed learning of form and definition knowledge assessed learning of meaning. Figure 2.2 shows mean accuracy per exposure condition for each task and indicates high accuracy levels across all of the tasks.

To compare performance between exposure conditions, mixed effects logistic regression models were computed separately for each task. Exposure condition (reading-only vs. reading-plus) was entered as fixed factor. For orthographic decision, the difference between exposure conditions was not significant (p = .513). TOWRE score was a significant covariate (p = .023), indicating that children with higher scores performed better at orthographic decision than children with lower scores. For Spelling, there was a significant difference between exposure conditions (p = .043), indicating greater accuracy in the reading-plus condition than the reading-only condition. This finding should be considered with caution, as performance on this task approached ceiling. TOWRE score was not a significant covariate (p = .071). Similarly, for definition knowledge, there was no significant difference between exposure conditions (p = .294), and TOWRE score was not a significant covariate (p = .282, see Appendix 2.5 for full models). In summary, there was good learning of both form and meaning, with no difference between reading-only and reading-plus conditions.

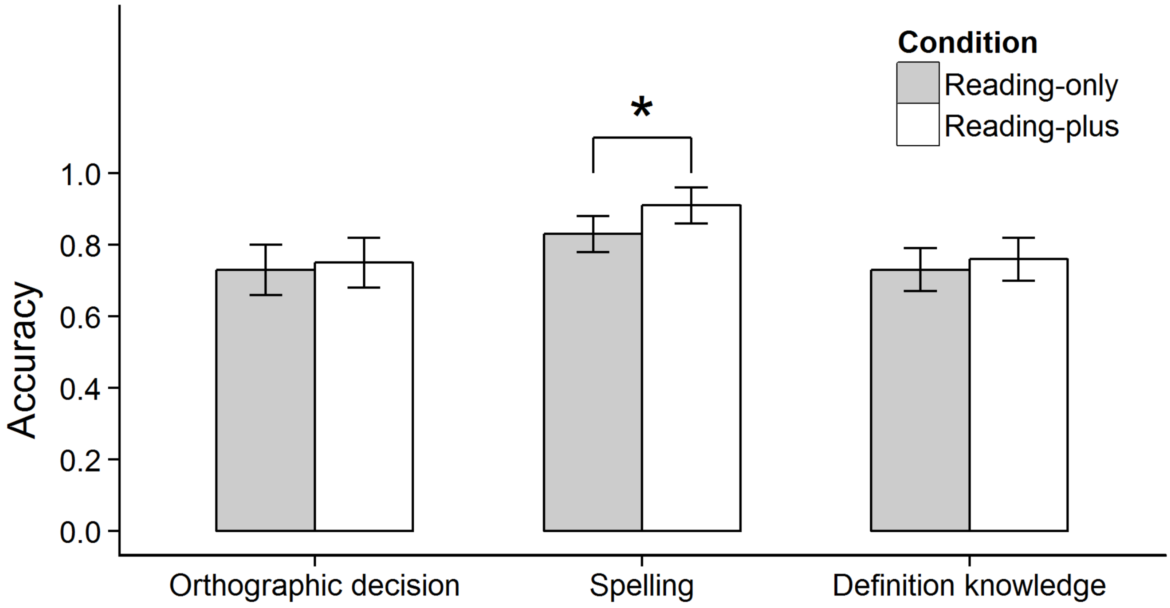


Figure 2.2. Mean accuracy scores for orthographic decision, spelling, and definition knowledge. Error bars show 95% CIs.

2.3.2 Did the novel words induce lexical competition effects for their neighbours?

Masked priming lexical decision. Due to an error in the experimental script, data from 18 trials (1.58%) had to be excluded from the analysis. Data were missing for one child in the reading-plus condition for the pre-exposure session. Overall, children performed with high accuracy on this task. Errors were made on only 76 of 1122 trials (6.77%). This was deemed insufficient for further analysis.

To index lexical competition between the novel words and their neighbour base words, priming effects on lexical decision were compared before and after exposure. Faster RTs for target base words primed by novel words (e.g. *wield* - *FIELD*) compared to unrelated primes (e.g. *quell* – *FRUIT*) would indicate facilitation by the novel words. Slower RTs for targets primed by novel words would indicate inhibition by the novel words. Only trials with correct responses were included in the analysis. Based on visual inspection of the qq-plot, RT values were inverse transformed to reduce positive skew, and one data point was removed. Figure 2.3 summarises the data per condition. For ease of interpretation, raw RT data are plotted. Inverse transformed data as used in the analyses are shown in Appendix 2.

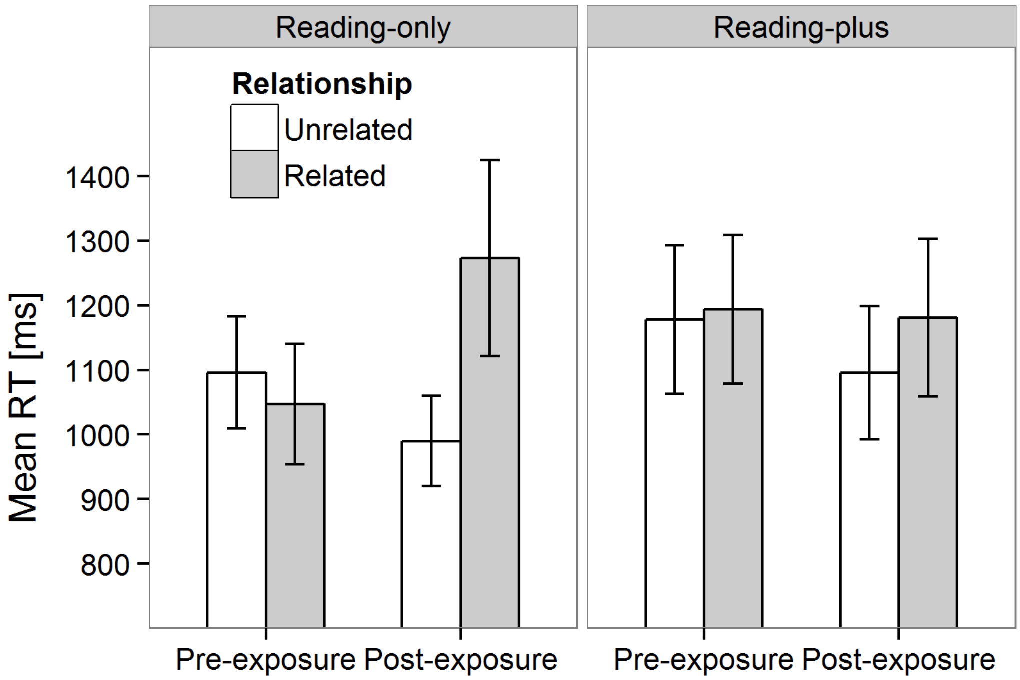


Figure 2.3. Mean RT data for masked priming lexical decision, split by exposure condition. Error bars show 95% CIs.

To analyse RT, a linear mixed effects model was fit with prime-target relationship (related vs. unrelated), test session (pre- vs. post-exposure) and exposure condition (reading-only vs. reading-plus) as fixed factors. TOWRE score was a significant covariate (t = -3.40), indicating that children with better word reading ability showed overall faster reaction times than those with lower word reading ability. The main effects of test session ($t < 0.01$) and exposure condition ($t = -0.04$) were not significant. There was a significant main effect of prime-target relationship ($t = 2.15$), indicating slower responses to targets preceded by related primes than unrelated primes. This main effect was qualified by a significant prime-target relationship x test session interaction ($t = 3.91$, Figure 2.4). There were no further significant interactions (see Appendix 2.5 for full models).

To explore the two-way interaction, pairwise comparisons were made by computing separate models for pre-exposure and post-exposure data. Before exposure, there was no difference between targets preceded by unrelated primes compared to related primes ($t = -1.11$), although numerically, raw RTs were shorter following related primes than unrelated primes. In contrast, after exposure children were significantly slower to respond to targets primed by related novel words than unrelated words ($t = 4.03$).

In summary, the interaction between test session and prime-target relationship indicated a switch in direction of priming effects, resulting in significant inhibitory priming of novel words on their neighbours after exposure. There was no difference in the magnitude or pattern of priming between reading-only and reading-plus conditions.

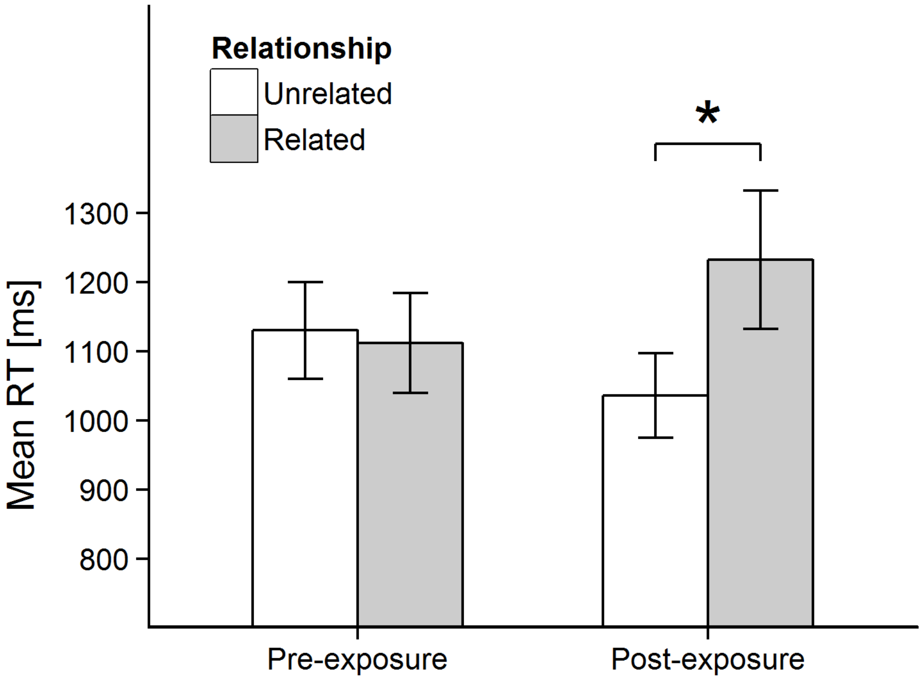


Figure 2.4. Mean RTs for the masked priming lexical decision task collapsed across exposure conditions, showing the interaction between prime-target relationship and test session. Error bars show 95% CIs.

Semantic relatedness. Data were missing from one child from the pre-exposure session and from one child for the post-exposure session, both from the reading-plus condition. Children performed with high accuracy on this task, with errors on only 82 of 1120 trials (7.32%). This was considered insufficient for further analysis. Based on visual inspection of the qq-plot, RTs on correct trials were log transformed and one data point was removed. Figure 2.5 summarises the data per exposure condition. For ease of interpretation, raw RTs are plotted. A linear mixed model was fit with base word status (neighbour of novel word trained in experiment vs. neighbour of control word), test session (pre- vs. post-exposure) and condition (reading-only vs. reading-plus) as fixed effects. The model with maximal model structure did not converge, so the random effects structure was simplified (see Appendix 2.5).

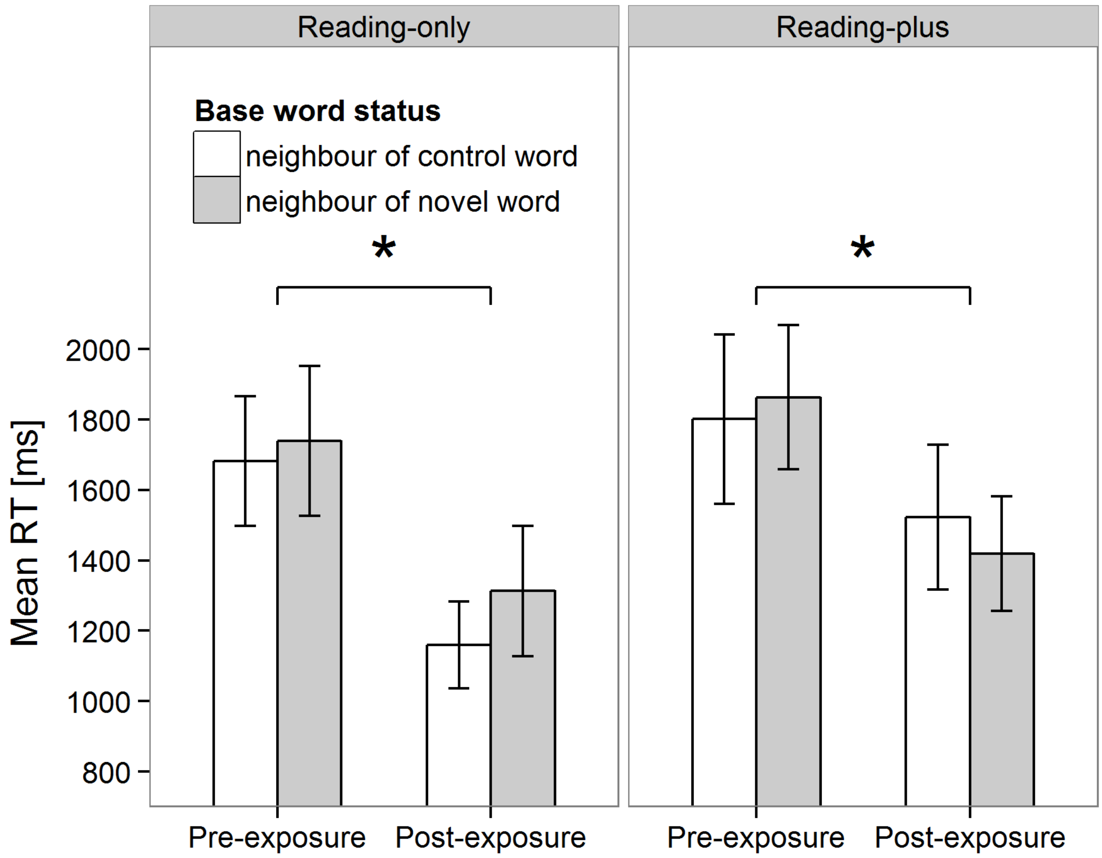


Figure 2.5. Mean RTs for the semantic relatedness task, split by exposure condition. Significance is indicated for the main effect of pre-exposure vs. post-exposure. Error bars show 95% CIs.

TOWRE was a significant covariate (t = -2.32), indicating that children with higher TOWRE scores shower faster RTs than children with lower TOWRE scores. The main effect of test session was significant (t = -10.82), showing faster RTs post-exposure (M = 1345, SD = 1236) compared to pre-exposure (M = 1766, SD = 1235). Neither the main effect of exposure condition (t = 0.99) nor the main effect of base word status were significant (t = -0.88), and there were no significant interactions (see

Appendix 2.5). In summary, children showed a general practice effect for both base words and control words, with no evidence of interference by novel words on their base word neighbours after training.

2.4 Discussion

The present experiment compared two types of measures of novel word learning in a sample of 9-11-year-old children. Children were assessed both in terms of explicit knowledge of the novel words they had experienced during exposure, measured using recognition and recall tasks, and novel word lexicalisation, measured using two lexical competition paradigms. Using these measures, the experiment compared learning and lexicalisation following two types of exposure conditions, reading-only and reading plus, which varied in the degree of directed focus on the novel words.

Children showed good knowledge of the novel words after three exposure sessions, as indicated by good spelling recognition on orthographic decision, good spelling recall, as well as good definition knowledge. There was also evidence of successful lexicalisation of novel words in the emergence of a prime-lexicality effect in the masked priming lexical decision task. In contrast to this, there was no evidence of interference by newly acquired words in the semantic relatedness task.

Performance on most tasks did not differ between children who read the stories aloud and those who completed additional, directed learning tasks on the novel words. Spelling performance was better in the reading-plus than the reading-only condition, but performance on this task approached ceiling. This provides support for the notion that reading meaningful text provides an opportunity for incidental orthographic learning that is equally successful as a more directed exposure paradigm.

2.4.1 Children's knowledge of novel word forms

Children's ability to recognise the correct spellings of the novel words was measured using orthographic decision. Performance was high on this task despite the use of stringent scoring criteria (acceptance of correct spelling and rejection of all three incorrect spellings) and was not affected by exposure condition, which was unexpected. Children who practised spelling the novel words during the exposure phase performed no better than children who just read them aloud, embedded in stories. This could partly be a result of the particular story characteristics, an issue discussed in more detail further on. In general, it is possible that the present task was easier than versions used in previous experiments, where children were exposed to nonwords (e.g. Nation et al., 2007). This is because the use of words made it impossible to create homophones that were always the same length as the correct spelling for some of the items (e.g. *propel* – *propell*). This may have made it easier for children to recognise the correct spelling than discriminating between homophone nonwords of the same length (e.g. *yait* – *yate*). The relative trade-offs of using words vs. nonwords are discussed in more detail in Chapter 6. Despite these methodological concerns, it is likely that performance accurately reflected good knowledge of the novel word forms, given that spelling performance was also high. Here, a group difference indicated that spelling benefitted from the reading-plus training, which involved copying the correct spelling repeatedly. However, performance generally headed towards ceiling. Spelling performance in the reading-plus condition could also have been inflated by the overlap in tasks used at exposure and test – both asking the children to spell the words. Using congruent tasks at exposure and test has previously been shown to result in better word learning than using incongruent tasks (Martin-Chang et al., 2007), in line with the encoding specificity principle (Tulving &

Thomas, 1973). Hence the impact of incidental vs. directed exposure on spelling warrants further investigation.

Children also showed good semantic learning of the novel words, indicated by their ability to provide definitions. Again, there was no difference between reading-only and reading-plus conditions. Here, the same concern applies regarding the nature of the story materials, with both conditions providing explicit, dictionary-style definitions. This may have reduced the effect of further directed exposure in the reading-plus condition. This limitation is discussed in more detail later in this chapter.

2.4.2 Lexical competition effects on neighbours of novel words

The present experiment assessed lexical competition effects using the prime-lexicality effect and semantic relatedness judgements. Lexical competition is considered a marker of lexical engagement, in contrast to orthographic decision, spelling, and definition knowledge, which only provide evidence of lexical configuration (Leach & Samuel, 2007; Chapters 1 and 3 of this thesis).

Regarding the prime-lexicality effect, this experiment is the first to demonstrate this effect following exposure to novel words in primary school children. This effect was indicated by a switch in direction of priming effects following exposure. Numerically, there was a facilitatory effect before exposure, with novel words priming faster responses to form-related targets, consistent with primes having nonword status to the children. However, this effect was not significant. This might be explained by relatively high variability in children's RTs paired with an insufficiently large sample[2]. However, after exposure, children showed significantly slower responses to base words when they were primed by trained words, compared to unrelated primes. This inhibitory

[2] This possibility was addressed by an analysis of pre-exposure data combined across experiments, where a significant facilitatory priming effect was found (see Chapter 6).

priming effect is consistent with lexical competition between trained words and their neighbours, a marker of lexicalisation. In line with the explicit measures of word learning, the prime-lexicality effect was not affected by exposure condition, indicating that both exposure conditions resulted in equally successful lexicalisation of novel words.

While all other measures pointed towards good word learning, there was no evidence that newly acquired words interfered with their neighbours in the semantic relatedness task. This was a novel task that had not previously been used to indicate lexicalisation. Interference from newly acquired words on a semantic task was found by Bowers et al. (2005), who used semantic categorisation to tap competition effects in adults. After exposure to nonwords such as *banara*, participants were slower to categorise *banana* as a fruit compared to before exposure. This task was quite different from the semantic relatedness task used in the present experiment. Due to the constraints of using words rather than nonwords, there was large variability in the semantic characteristics of the target words, e.g. part of speech, semantic category and imageability. This could explain why the present semantic relatedness task was insensitive to the novel word lexicalisation evident in the prime-lexicality effect.

2.4.3 Limitations and future research

These results are in line with previous research showing that children are able to acquire novel vocabulary via incidental exposure (reviewed in Swanborn & DeGlopper, 1999). However, there were some characteristics of the reading materials which limit the extent to which the reading-only condition is comparable with children's incidental learning via everyday reading experiences. Given that the prime-lexicality effect as a measure of lexicalisation has only been demonstrated after explicit training and only in adults (Qiao & Forster, 2013), the present reading materials were designed to maximise word learning, while still maintaining the framework of fictional stories. As a consequence, several factors served to artificially increase the salience of the novel words. First, each novel word appeared six times in a relatively short text, leading to a high ratio of novel words to context. A higher ratio of novel words to context words has been shown to predict higher rates of acquisition in a meta-analysis of incidental vocabulary learning (Swanborn & DeGlopper, 1999). Second, each story was paired with an illustration that always related to the novel words. Third, the premise of the stories was related to word learning, as this was the task faced by the protagonists in the narrative. Finally, the stories contained explicit definitions of the target words. As a consequence, one could argue that the reading-only condition was not representative of incidental learning, as it happens in natural reading experience. Indeed, our materials would not meet some of the stricter definitions of incidental learning, e.g. that 'context is not intentionally made more transparent than it would be in natural text' (Swanborn & DeGlopper, 1999, p. 264). Thus, while the present results are encouraging regarding the potential for children to learn novel word forms from stories, it remains unknown whether incidental learning as it happens in more natural reading experience would result in equally successful word learning and lexicalisation as found here. This issue was addressed by the following experiments, which used a new set of materials addressing the limitations discussed here.

In conclusion, with the exception of spelling, Experiment 1 found no benefit of the directed exposure to novel words offered by the reading-plus condition relative to the more incidental reading-only condition, either for children's knowledge of the words or the extent to which the novel words competed with their orthographic neighbours This confirms that encountering words during reading provides a powerful opportunity to learn about form and meaning. While learning was clear, the extent to which it reflects incidental learning remains an open question, due to some limitations of the reading materials used. In general, however, lexical competition as indexed by the prime-lexicality effect showed good promise as a sensitive measure of novel word form learning. In subsequent experiments, it was used to investigate a number of factors of interest to further elucidate the nature and mechanisms of orthographic learning in children. Several specific theoretical questions could be raised on the basis of Experiment 1. Good learning and lexicalisation was evident after an extended exposure paradigm of three sessions totalling 18 exposures per word – could this be found after fewer exposures (Experiment 2)? How does learning via meaningful stories compare with exposure without semantics (Experiments 3.1 and 3.2)? And finally, how does repetition of the same stories as used here compare to experiencing the word in different and more diverse stories (Experiment 4)?

Chapter 3:

Effects of Exposure on Novel Word Learning and Lexicalisation

Experiment 2

3.1 Introduction

Exposure is a basic requirement for learning about the orthography of novel word forms. This chapter concerns the question of how much exposure to novel word forms is necessary for their acquisition by children– specifically when they are encountered incidentally in meaningful text. Existing research has revealed differences in the time course of orthographic learning depending on how it was measured. In particular, two types of measures can be contrasted. When assessed using recall or recognition of the novel words' spelling, children require relatively few exposures to show evidence of orthographic learning (e.g. Nation et al., 2007; Share, 1999). In contrast to this, when lexical competition between novel word forms and their neighbours was used to indicate learning in adults (e.g. Qiao & Forster, 2013), lengthier training involving more exposures than those used in children's orthographic learning studies was required. It has been argued that these two types of measures map onto distinct stages of word acquisition (lexical configuration vs. lexical engagement), which rely on different memory systems (temporary episodic memory traces vs. stable, integrated memories) with different time courses (fast vs. slow) (e.g. Bakker et al., 2014; Leach & Samuel, 2007). So far, these distinctions have been discussed in the literature on spoken word recognition, in children and adults (e.g. Henderson et al., 2012; Leach & Samuel, 2007), and written word learning in adults (e.g. Qiao & Forster, 2013), but not yet in developing readers. Experiment 2 was designed to extend existing findings to children's orthographic learning via incidental story reading, comparing the effects of amount of exposure on two types of learning measures in the same experiment.

3.1.1 The time course of children's orthographic learning

Several studies have investigated effects of number of exposures on children's orthographic learning. Share (2004) exposed 3rd graders learning to read Hebrew to pseudowords either once, twice, or four times within a single session. Target words were embedded in short texts which children read aloud without assistance. Spelling and orthographic choice performance demonstrated reliable orthographic learning that did not differ significantly based on the number of exposures. In other words, a single exposure was sufficient for good orthographic learning. There was also no effect of test delay – children performed equally regardless of whether tested after three days, one week, or one month after exposure. This suggests that in Hebrew, lasting orthographic learning can be found after a single exposure. De Jong and Share (2007) investigated orthographic learning in Dutch 3rd graders (aged 8 to 9 years), comparing two or six exposures per novel word. Children either read stories silently or aloud. In line with Share's (2004) findings, orthographic learning measured by orthographic choice, naming time and spelling was equally good following two or six exposures, for both silent and oral reading.

Findings from studies conducted with children learning to read in English show somewhat different results. Bowey and Muller (2005) exposed 3rd graders (aged 7 to 9 years) to novel words either four or eight times. In this experiment, children were reading silently. The results revealed significantly better learning after eight than four exposures, as indicated by orthographic choice, both immediately after exposure and after a six-day delay. Performance was on the whole weaker after six days than immediately after exposure. Nation et al. (2007) also reported an effect of number of exposures and test delay. Children from grades 3 and 4 were exposed to novel words either once, twice, or four times. Children read aloud target words either individually from flashcards or embedded in stories, which did not affect performance. Children were tested after one day and again after seven days. Children showed some evidence of learning after a single exposure, but performed significantly better at orthographic choice after four exposures than after one. This was the case at both test times, but performance was generally better after one day than after seven. In summary, both experiments showed that children reading in English benefitted from additional exposures, and that learning was not as robust over time as in Hebrew. The discrepant results across Hebrew, Dutch and English might be due to differences in orthographic depth. Hebrew and Dutch are shallow orthographies that are thus easy to decode, whereas English is a deep orthography with lower levels of consistency between graphemes and sounds. This might make orthographic learning in English a more difficult task than in Hebrew or Dutch. However, despite the main effects of exposure found in English, it is important to note that even ‘high exposure’ conditions involved relatively few exposures (eight in Bowey and Muller, 2005; four in Nation et al., 2007), and some orthographic learning was evident after a single exposure session.

3.1.2 The time course of lexical competition effects in skilled visual word recognition

The studies described above manipulated the number of exposures within a single learning session, rather than comparing multiple numbers of sessions; a single exposure session was sufficient to produce orthographic learning. In contrast to this, competition effects as indicators of lexicalisation of novel word forms have so far not been found less than 24 hours after a single training session (Bakker et al., 2014; Qiao et al., 2009; Qiao & Forster, 2013). Across studies, training also generally involved more exposures per novel word per session than in the above children's studies. For example, regarding the prime-lexicality effect, in Qiao et al. (2009), participants completed ten blocks of typing the novel words and they also read through the entire list of novel words at the end. This was insufficient to produce inhibitory priming immediately after training. Qiao and Forster (2013, Experiment 1) trained participants during a single session involving pictures and definitions for each novel word, with seven exposures to each written form. This was also insufficient to produce a prime-lexicality effect. Experiment 2 (Qiao & Forster, 2013) used the same training tasks of Experiment 1, but repeated over four sessions, resulting in four times the number of exposures than Experiment 1. This time a prime-lexicality effect did emerge. This suggests that multiple training sessions might be necessary for successful lexicalisation, although this may differ depending on how the words are trained. For instance, more exposures might be necessary when using a training paradigm that is less directed towards word learning and less semantically rich than Qiao and Forster's (2013).

Bakker et al. (2014, Experiment 3) used a different training regime and a different measure of lexical competition, but found broadly compatible results. As in Qiao and Forster (2013), they trained adults on nonwords that were the only neighbours of existing base words, but the experiment was conducted in Dutch. Training involved a letter-monitoring task, asking participants to indicate whether varying target letters were present in the novel words. As such, training drew attention towards the individual letters of the novel words, but did not provide any semantic information. Competition effects were assessed immediately after training, after 24 hours, and after eight days. Competition between the training words and base words was measured using a semantic categorisation task, based on the one developed by Bowers et al. (2005). Participants were asked to categorise base words as artefacts or natural objects. Slower responses to base words with trained neighbours compared to those with untrained neighbours would be evidence of lexical competition. Competition effects did not emerge immediately after training, but reached significance (by items and subjects) after 24 hours, and remained so (by subjects) on day eight. Taken together, Qiao and Forster (2013) and Bakker et al. (2014) both failed to observe competition effects between written words immediately after a single training session; instead, competition effects took at least 24 hours to emerge.

In summary, children show fast orthographic learning when learning is measured by orthographic choice or spelling (de Jong & Share, 2007; Nation et al., 2007; Share, 2004), while the adult studies using competition paradigms only found evidence of lexicalisation after more extended training (Bakker et al., 2014; Qiao & Forster, 2013). This dissociation in time courses based on how orthographic learning was measured maps well onto theoretical distinctions proposed both at the level of cognitive word learning stages, and the neural memory systems that support learning. These will be discussed shortly, but it is of course important to note that the dissociation in time course could also be due to differences between individual experiments other than the way learning was measured. For example, learning may have been comparatively slower or faster due to differences in samples (children vs. adults), languages, or differences in the particular exposure paradigms chosen. The present experiment controlled for some potential confounding factors by measuring orthographic decision (an improved variant of orthographic choice), spelling, and lexical competition in the same children, keeping the exposure paradigm constant throughout.

3.1.3 Lexical configuration vs. lexical engagement

At the cognitive level, it has been suggested that orthographic decision and spelling map on onto a different stage of word learning than lexical competition. This is reflected in the different ways that learning can be inferred from these tasks. While orthographic decision and spelling measure performance on the trained word, lexical competition paradigms such as the prime-lexicality effect measure learning via the effect of a newly learned word on its orthographic neighbour; that is, learning is inferred indirectly via an interference or competition effect. As such, lexical competition is argued to reflect a lexical representation of the novel form that is fully integrated into the lexicon. Orthographic choice and spelling reflect knowledge of the novel form, but integration into the lexicon is not a direct prerequisite.

The presence or absence of integration into the lexicon is at the heart of a distinction proposed by Leach and Samuel (2007) between lexical configuration and lexical engagement. Lexical configuration refers to factual knowledge of lexical properties, e.g. a word's phonology, meaning and spelling, and is tapped by tasks such as orthographic decision and spelling. Lexical engagement refers to a word's interaction with other entries in the lexicon, e.g. as tapped by lexical competition as indexed by the prime-lexicality effect. In the spoken word recognition literature, it has been shown that more exposures are needed to support lexical engagement than lexical configuration (e.g. Gaskell & Dumay, 2003). As Leach and Samuel summarise (2007, p. 340), 'it is one thing to learn the form of a new word, and something else for the representation to take on all the "rights and responsibilities" of a fully functional lexicon'. While this distinction originates from the spoken language literature, Qiao and Forster (2013) have since applied it to the learning of novel written words by adults. Experiment 2 of this thesis was the first to investigate whether these principles also apply to children's orthographic learning of novel words.

At the level of memory systems, the distinction between lexical configuration and lexical engagement has in turn been argued to fit the Complementary Learning Systems (CLS) framework (McClelland et al., 1995). This framework proposes a fast learning system based on temporary episodic memories encoded in the hippocampus, and a slow learning system which integrates these memories into a distributed memory network based in the neocortex. The transition from episodic memories to memories integrated with existing memories has been shown to rely on consolidation processes, often during sleep (e.g. Davis & Gaskell, 2009). With regard to orthographic learning, one could argue that episodic memories of novel word forms could be created quickly and serve lexical configuration. This fits with the evidence of rapid orthographic learning when novel words have to be recognised or recalled (de Jong & Share, 2007; Nation et al., 2007; Share, 1999), which could be achieved on the basis of episodic 'novel word memories' (Bakker et al., 2014, p. 117). A comparatively lengthier consolidation process may be required for the integration of such memories into knowledge of existing words to allow lexical engagement, which fits with the finding that novel written words only exert lexical competition effects after extended training (e.g. Qiao & Forster, 2013).

To summarise the relationship between cognitive processes, memory systems and experimental tasks, episodic 'novel word memories' have been argued to be sufficient to serve lexical configuration, i.e. lexical knowledge of new words, such as their sound, form or meaning. At the task level, lexical configuration can be tapped using recall or recognition tasks of the novel words themselves. On the other hand, lexical engagement, e.g., interactive processes between novel and existing word entries in the lexicon, has been argued to rely on consolidated, integrated lexical representations that are slower to emerge. At the task level, lexical engagement by novel written words can be measured using the prime-lexicality effect to infer lexical competition.

3.1.4 The nature of the exposure paradigm

A final theoretical issue relevant to the present experiment is the nature of the exposure paradigm. None of the studies described above used exposure paradigms designed to model incidental learning akin to children's reading experiences in real life. Consequently, they may be informative regarding word learning under specific laboratory conditions, but may not generalise to incidental learning via reading stories in children, which remains the focus of this thesis. Studies on orthographic learning made use of story contexts, but these were relatively short with averages of 126 words (Share, 1999, Experiment 1), 127 words (Bowey & Muller, 2005) and 94 words (Nation et al., 2007). The resulting high ratio of novel words to context, in addition to the use of artificial nonwords, may have drawn specific attention to the novel words. Furthermore, explicit definitions of the novel words appeared to be given as part of the stories (e.g. 'the hottest town in the world', Share, 1999). The present experiment sought to engage children in an exposure paradigm more akin to real-life incidental learning. Story contexts were improved relative to those used in Experiment 1: explicit definitions of the novel words were removed, the premise of the stories was no longer related to learning about words, stories were increased in length to an average of 162 words, and the number of repetitions per novel word per story was reduced from six to four. The experiment constituted a first investigation of the effects of number of exposures on children's orthographic learning using measures of both lexical configuration and lexical engagement, following an incidental learning paradigm using words rather than artificial nonwords.

3.1.5 The present experiment

The present experiment focussed on the time course of orthographic learning as indexed by spelling and orthographic decision performance on the novel words themselves, and the emergence of lexical competition effects on their orthographic neighbours – designed to contrast lexical configuration and lexical engagement. Specifically, the experiment manipulated the number of exposures to investigate how many exposures are necessary for good learning to emerge on the two types of measures. In Experiment 1, good performance on both lexical configuration and lexical engagement tasks was evident after 18 total exposures spread over three exposure sessions. In the present experiment, children either encountered novel words in a single session with four exposures, or in three sessions with four exposures each, resulting in 12 exposures in total.

Based on previous research, I predicted lexical configuration (as measured by orthographic decision and spelling) would emerge after fewer exposures than lexical engagement of these words with their neighbours (as measured by the prime-lexicality effect). More generally, Experiment 2 also served to replicate the finding of a prime-lexicality effect seen in Experiment 1, using different materials and a different sample. A new set of story contexts was created, intended to resemble children's real-life incidental learning more closely. Finally, children's learning of the novel word meanings was also assessed via definition knowledge, to measure the emergence of semantic knowledge. It was anticipated that definition knowledge would be better after a greater number of exposures, given that this task was relatively demanding.

3.2 Method

3.2.1 Design

The experiment involved a mixed design with a between-subjects manipulation of exposure (four exposures vs. twelve exposures). The experiment consisted of three phases – pre-exposure baseline, exposure phase, and post-exposure test. Data for the masked priming lexical decision task were collected before and after exposure, yielding a within-subject factor of test session (pre-exposure vs. post-exposure). Orthographic decision, spelling and definition knowledge data were collected once, after exposure.

3.2.2 Participants

Twenty-two children (12 female, 10 male) from Year 5 and Year 6 classes participated in the experiment. Children were selected on the basis of being monolingual native English speakers scoring no more than 1*SD* below the standardised average on the TOWRE (Torgesen at al., 1999). This test consists of two subtests – word reading and nonword reading. Performance on the two subtests was highly correlated (.78). Children in the four-exposures condition scored a little lower on the TOWRE than the twelve-exposures group, so data from two children with the lowest scores were excluded, leaving ten children in each condition[3]. Descriptive characteristics of the final samples are shown in Table 3.1. Raw TOWRE scores indicate reading ability independent of age, which was used in subsequent analyses. Standard scores reflect reading ability standardised for age and are reported here to illustrate the nature of sample relative to the general population. The overall mean total standard score was 110.6 (*SD* = 9.11), indicating that children performed at the upper

[3] For the twelve-exposures condition, data from the 10 children collected as part of Experiment 4 was used. For the single-session condition, data from 12 additional children was collected.

end of the population average. Groups did not differ significantly in age or reading ability (see Table 3.1). Children were recruited using an opt-out procedure, where parents were sent two information sheets over the course of two weeks with the opportunity to opt out of their child's participation in the study by returning a response slip.

Table 3.1. Descriptive characteristics of four-exposure and twelve-exposure groups

	Four exposures (*N* = 10)		Twelve exposures (*N* = 10)		Group comparison
Measures	*M*	*SD*	*M*	*SD*	*p*
Age (years)	10.16	.77	10.68	.83	.166
TOWRE total raw score[a]	111.4	7.38	117.5	17.19	.316
TOWRE total standard score[b]	110.3	6.09	110.9	11.74	.888

Notes. [a] total number of words read correctly on both subtests in 45 seconds per test, [b] M = 100, SD = 15.

3.2.3 Materials and procedure

3.2.3.1 Word stimuli

The 20 word pairs were identical to those used in Experiment 1, comprising novel words and their orthographic neighbours (base words). For counterbalancing purposes, word pairs were once again divided into lists A and B. Children were allocated to either list based on the order in which they participated in the experiment.

3.2.3.2 Pre-exposure baseline assessment of lexical engagement

Masked priming lexical decision was used to assess lexical engagement between novel words and their orthographic neighbours. For Experiments 2, 3.1, 3.2 and 4, the task was presented on a Dell Latitude E6530 laptop using E-Prime 2.0 and a Serial Response Box (Psychology Software Tools, Inc.) was used to record responses instead of a keyboard. Otherwise the task was identical to that used Experiment 1. Priming effects were compared before and after exposure in order to measure the emergence of lexical competition. Prior to exposure, novel words were expected not be lexicalised and thus behave like nonwords, showing facilitatory priming effects on target neighbour base words relative to unrelated primes. After exposure to the novel words, a relative reduction of this facilitatory effect, resulting either in the absence of priming or the emergence of an inhibitory effect, would provide evidence of lexical competition driven by lexicalisation of the novel words. Due to scheduling constraints, some children completed the pre-exposure assessment on the same day as the start of the exposure phase, whereas others completed the task on a separate day before the exposure phase.

3.2.3.3 Exposure phase

3.2.3.3.1 Story contexts. Twenty stories were created, one per novel word (mean length = 162 words), divided into lists A and B. Stories did not differ significantly in length between the two lists (p = .478). One novel word appeared per story, with four repetitions within each story. Stories did not provide explicit definitions of the novel words (cf. Experiment 1), but made it possible to infer the meanings from the context. The stories did not follow on from one another, but as in Experiment1, they did involve a consistent set of characters. Stories were laid out on A5 pages, modelling the style and format of children’s fiction. Full stories can be found in Appendix 3.1.

3.2.3.3.2 Exposure conditions

Twelve-exposures condition. The twelve-exposures condition involved three sessions spread over different days with four exposures per novel word per session (12 exposures in total). Exposure sessions were not always completed on consecutive days, due to scheduling constraints. The number of days across which the exposure phase was completed varied from 3 to 16 days (M = 6.7, SD = 4.57). In session 1, children were told they would be reading a story printed on several cards similar to chapters in a book. Prior to reading the stories, children were shown an A4 sheet introducing the four main characters of the stories, members of a group called the 'Caterpillar Club', and their rivals, the 'Gloopy Goo Gang'. The card contained a coloured illustration of each character and two descriptive sentences each (see Appendix 3.1). They were told they would be reading short stories about these characters. They were asked to read the stories aloud to the experimenter, in random order. After half of the stories, children engaged in a non-verbal filler task. In sessions 2 and 3, children read the stories aloud once more. Children were corrected if they read out the novel words incorrectly and were generally assisted in their reading as required.

Four-exposures condition. The four-exposures condition involved a single session with 4 exposures per target word (4 exposures in total). The procedure was identical to session 1 of the twelve-exposures condition.

3.2.3.4 Post-exposure assessments

These assessments were administered in the order listed below. Children were tested between 1 and 5 days after completing the exposure phase (M = 1.85, SD = 1.39), depending on their availability. The test delay did not differ between groups (p = .641).

Word definition recall. To assess lexical configuration of novel word meanings, children were asked to verbally define each novel word as shown printed on cardboard cards.

Masked priming lexical decision. To assess lexical engagement**,** children completed the lexical decision task as before, without the practice version.

Orthographic decision. This was used to assess lexical configuration of novel word forms, in particular the ability to recognise correct spellings. The task was a modification of the one used in Experiment 1, involving the correct spelling, a homophone foil, and two one-letter-different foils per novel word. Stimuli were shown on cardboard cards rather than a computerised task, to stop children from rushing through the task without paying close attention. To indicate their orthographic decision, children were asked to sort the cards into a two stacks – correctly and incorrectly spelt words. They were told to look carefully and take their time. The task was scored based on Wang et al.'s (2013) procedure – for each novel word, performance was coded as accurate if a Yes response was made to the correct spelling and all three incorrect spellings were rejected. This is a relatively stringent way to assess performance on this task.

Spelling. This was used to assess lexical configuration of novel word forms, in particular the ability to produce correct spellings. The trained novel words were dictated in a random order and children were asked to spell them. Responses were scored as correct or incorrect.

3.3 Results

Analyses were performed using linear mixed effects models following the same principles outlined in Chapter 2. For models with continuous outcome variables, significance was determined using the cut-off point of $|t| > 2$, based on recommendations by Barr et al. (2012). For models with binary outcome variables, mixed effects logistic regression models were computed. Here, the cut-off point for significance was $p < .05$. To control for individual variation in word reading ability, children's total raw TOWRE scores were included as covariates in the analyses.

3.3.1 Effects of exposure on lexical configuration of novel words

Three tasks assessed lexical configuration of novel words. Of these, two tasks assessed knowledge of novel words forms, namely orthographic decision and spelling. The definition knowledge task assessed knowledge of novel word meanings. Figure 3.1 shows mean accuracy per exposure condition for each task. To compare performance between exposure conditions, mixed effects logistic regression models were computed for each task. Exposure condition (four exposures vs. twelve exposures) was entered as fixed factor. For orthographic decision, data were missing from one child in the four-exposures condition. Both maximal and simplified models with TOWRE score entered as a covariate failed to converge, so TOWRE score was not entered for the analysis of orthographic decision. The difference between conditions was not significant ($p = .224$). For spelling, there was also no significant difference between conditions ($p = .762$). TOWRE score was a significant covariate ($p = .0496$), indicating better spelling performance for children with higher TOWRE scores than children with lower scores. For definition knowledge, there was a significant difference between the four-exposures and the twelve-exposures conditions ($p < .001$). Children in the twelve-exposures condition correctly defined significantly more novel words than those in the four-exposures condition. TOWRE score was not a significant covariate (*p* = .329). See Appendix 3.2 for full models.

In summary, the number of exposures did not affect learning of the novel words forms, but children showed better semantic learning after twelve exposures than after four exposures.

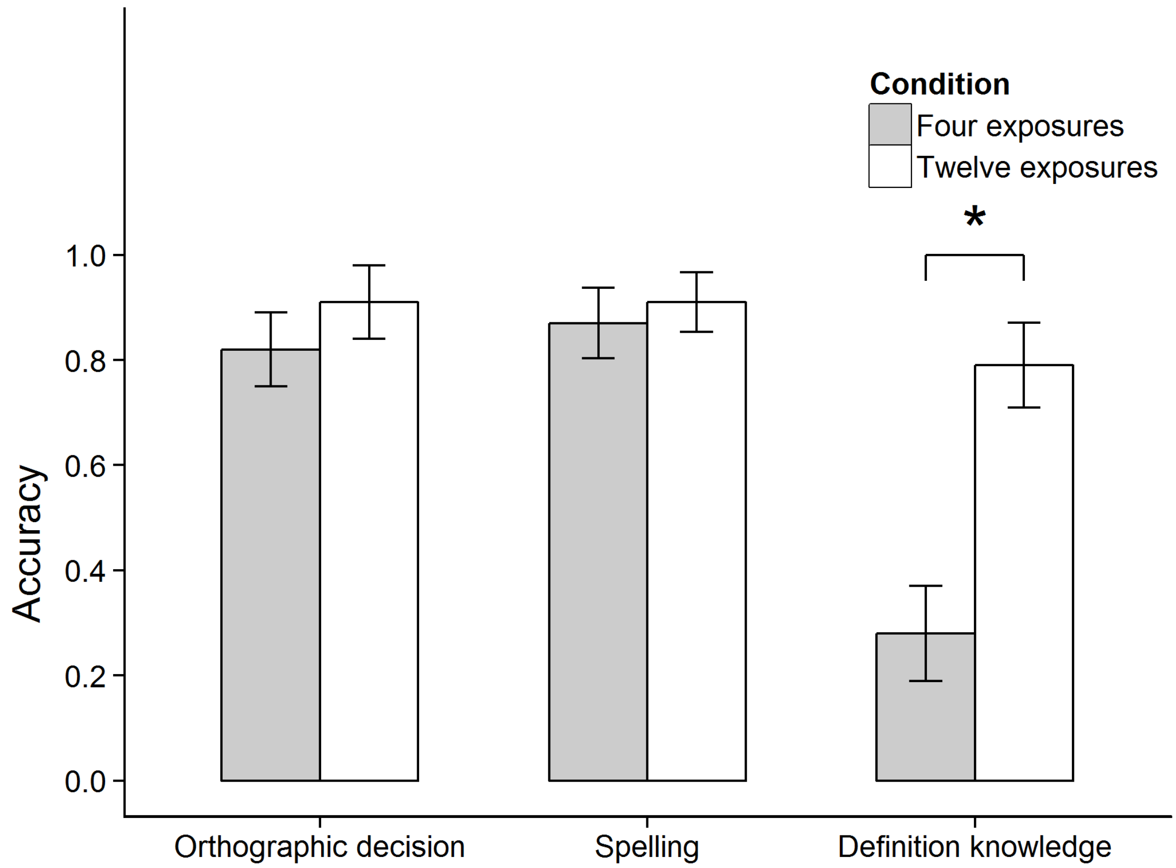


Figure 3.1. Mean accuracy scores for orthographic decision, spelling, and definition knowledge comparing exposure conditions. Error bars show 95% CIs.

3.3.2 Effects of exposure on lexical engagement of novel words with their neighbours

Masked priming lexical decision. Due to an error in the experimental script, data from 36 (4.5%) trials had to be excluded from the analysis. Children performed with high accuracy on this task. Errors were made on 40 out of 764 (5.34%) trials across sessions and this was deemed insufficient for further analysis. Only trials with correct responses were included in the RT analysis. Based on visual inspection of the qq-plot, inverse RT values were used in the analysis to reduce positive skew, and one data point was removed. For ease of interpretation, raw RT data are plotted. Figure 3.2 summaries the data per condition. Inverse transformed data as used in the analyses are shown in Appendix 3.

Linear mixed effects models were computed with prime-target relationship (related vs. unrelated), test session (pre- vs. post-exposure) and exposure condition (four exposures vs. twelve exposures) as fixed factors. TOWRE score was a significant covariate (t = -6.65) indicating that children with higher TOWRE scores showed overall faster reaction times than those with lower scores. The main effects of prime-target relationship (t = -0.92), test session (t = -0.97), and exposure condition (t = 0.61) were not significant. There was a significant test session x exposure condition interaction (t = -2.24). This interaction was further qualified by a significant three-way interaction between prime-target relationship, test session and exposure condition (t = -2.10). There were no further significant interactions.

The three-way interaction was explored by computing separate models for each exposure condition (see Appendix 3.2 for full models). TOWRE score was a significant covariate both in the four-exposures condition (t = -4.53) and the twelve-exposures condition (t = -6.36), indicating that children with higher TOWRE scores showed overall faster reaction times than those with lower scores. In the four-exposures condition, there was a significant main effect of test session (t = -2.70), indicating faster reaction times after exposure than before exposure. The main effect of prime-target relationship was not significant (t = -0.57), and there were no significant interactions (See Appendix 3.2). In the twelve-exposures condition, neither the main effects of test session (t = 0.86) nor prime-target relationship (t = -0.54) were significant. There was however a significant prime-target relationship x test session interaction (t = 2.52). Pairwise comparisons revealed a marginally significant effect of prime-target relationship before exposure (t = -1.8), indicating a trend towards faster RTs to targets primed by related than unrelated words. After exposure, the effect of prime-target relationship was not significant (t = 1.2). In summary, for the twelve-exposures condition only, the interaction between test session and prime-target relationship indicated a switch in direction of priming effects from nonword to word-like priming behaviour. This prime-lexicality effect did not emerge after four exposures.

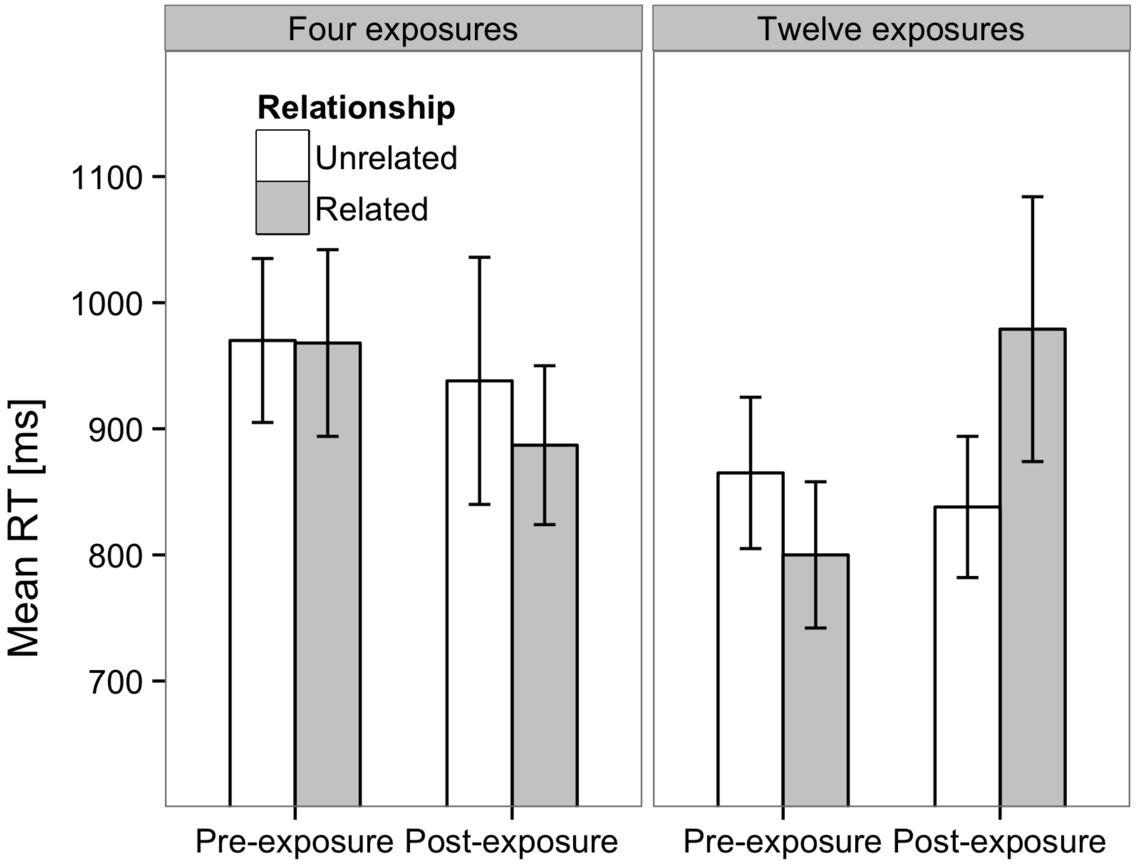


*Figure 3.2.*Mean RTs on the masked priming lexical decision task, split by exposure condition. Error bars show 95% CIs.

3.4 Discussion

This is the first experiment to examine the effects of number of exposures on measures both of lexical configuration and lexical engagement of novel word forms in children, using an incidental-learning paradigm. With regard to lexical configuration, children who were exposed four times in a single session to novel words showed equally good spelling recognition and recall as children exposed twelve times over three sessions. In contrast to form learning, recall of novel word meanings was significantly better after twelve exposures than after four exposures. With regard to lexical engagement, the time course observed differed from the above observation of rapid orthographic learning; only after twelve exposures over three sessions did the novel words exhibit lexical competition effects on their orthographic neighbours. Each of these findings will be discussed in turn.

3.4.1 Effects of number of exposures on lexical configuration

The present experiment assessed lexical configuration of novel word forms using recognition (orthographic decision) and recall (spelling). On both tasks, performance was equally good for children who experienced four exposures in a single session and children who experienced twelve exposures over three sessions. This is consistent with findings showing orthographic learning after relatively few exposures, and in a single session, in the self-teaching literature (de Jong & Share, 2007; Nation et al., 2007; Share, 1999). Studies in English nonetheless reported an effect of number of exposures, unlike the present experiment (Bowey & Muller, 2005; Nation et al., 2007). Bowey and Muller (2005) found significantly better performance on orthographic choice after eight exposures compared to four exposures, while the present experiment found no difference between four and twelve exposures. This discrepancy could be due to a difference in age, as the children in Bowey and Muller (2005) were around two years younger than in the present experiment. Orthographic learning has been shown to improve throughout primary school (Share, 1999). It could also be explained by differences in the way children were trained, as Bowey and Muller (2005) asked children to read silently, while in the present study, children read aloud and received feedback on their pronunciation. This provision of consistent lexical phonology for the novel words may have allowed them to acquire their spellings more quickly than during silent reading. The age difference also applies to Nation et al. (2007), who investigated children around one year younger than the present experiment. They also found an effect of exposure comparing one, two and four exposures, i.e. lower numbers of exposure than the present experiment. Potentially then, an effect of exposure on lexical configuration might be found for incidental learning of written words if exposures were thus reduced.

While children demonstrated good knowledge of word spellings after four exposures, learning about their meanings (as assessed by definition knowledge) was affected by the amount of exposures received. Children produced significantly more correct definitions for the novel words after twelve exposures over three sessions than after four exposures in a single session. This suggests that in the present study, lexical configuration of word meanings took longer to emerge than lexical configuration of word forms. One should however be cautious in using these results for a direct comparison of lexical configuration of form vs. meaning per se, for which the present experiment was not intended. Learning of word meanings may have taken longer than learning of spellings because of the nature of the incidental learning paradigm chosen here. In order to acquire semantic knowledge of the novel words, children had to infer their meanings from story contexts. Good performance at definition knowledge could not be achieved simply recalling information in the same form that it was given at exposure, as children had to construct their own definitions. On the other hand, novel word spellings were available in the stories without the need for further inferences, and no information beyond what they had encountered during story reading was required.

3.4.2 Effects of number of exposures on lexical engagement

While the number of exposures did not affect orthographic learning as assessed by the lexical configuration measures, exposure did affect the emergence of lexical engagement, as measured by the prime-lexicality effect. After four exposures in a single session, there was no difference in the priming effects of the novel words on their neighbours (calculated by comparing RTs between related and unrelated primes) compared to before exposure. After twelve exposures, priming effects were significantly different after exposure compared to before exposure, as confirmed by a significant interaction between prime-target relationship and test session. Before exposure, there was marginally significant facilitation when novel words primed their neighbours, consistent with them behaving like nonwords. The size of the sample (ten children in twelve-exposures group) likely contributed to the lack of a significant effect[4]. After exposure, there was no evidence of facilitation, which is consistent with the emergence of lexical competition between the newly lexicalised novel words and their neighbours.

These findings are consistent with Qiao and Forster (2013), who found evidence of lexical competition (indicated by an absence of facilitation by novel words) after four sessions, but not immediately after a single exposure session. While participants were tested immediately after the single exposure session in Qiao and Forster (2013, Experiment 1), in the present experiment, children were tested 24 hours after exposure, allowing a period of sleep and therefore potential for offline consolidation. Nonetheless, the inclusion of this consolidation period did not result in significant competition after a single session. In contrast, Bakker et al. (2014) found competition effects on a semantic categorisation task in adults 24 hours after a single session. Importantly, their training involved monitoring letters contained in the training words, while the present paradigm was based on reading aloud words in meaningful context. It remains to be seen whether lexical competition effects can be found 24 hours after a single exposure session in children if the exposure paradigm draws specific attention to individual letters of the novel words, rather than via incidental exposure. The use of a semantic categorisation

---

[4] This possibility was addressed by an analysis of pre-exposure data combined across experiments, where a significant facilitatory priming effect was found (see Chapter 6).

task rather than the prime-lexicality effect is also noteworthy, as it has been suggested that competition effects observed on the former could be at least partially driven by interference of episodic memories of the novel words (see Qiao et al., 2009; Qiao & Forster, 2013, for further discussion). If this was the case, it could explain why Bakker et al. (2014) were able to observe effects after a single exposure session. However, the fact that effects only emerged after a 24-hour delay involving sleep and not immediately after exposure suggests that consolidation-dependent lexicalisation likely contributed to the effect, rather than episodic memory alone. The issue of lexical competition vs. episodic interference is discussed in greater detail in Chapter 6.

The emergence of a prime-lexicality effect after twelve exposures constitutes a replication of the effect seen in Experiment 1 in a new sample of children and using different materials. However, lexical competition observed in the present experiment on the whole appeared weaker than in Experiment 1. In Experiment 1, the priming effect was inhibitory after exposure both for children reading stories (reading-only) and children completing directed activities (reading-plus). In Experiment 2, lexical competition was indicated by an absence of facilitatory priming rather than inhibitory priming. This might be explained by insufficient power to detect inhibition in Experiment 2 (10 children in twelve-exposures group in Experiment 2 compared to 31 children collapsed across groups in Experiment 1). Alternatively, the difference between experiments might constitute a further effect of number of exposures, with stronger lexical competition after 18 exposure in Experiment 1 than after 12 exposures in Experiment 2. However, this cannot be concluded for certain, as the difference in the strength of lexical competition could also be a result of the differences in the story materials used. Experiment 1 used an exposure paradigm that could be described as ‘less incidental’ than the present one, given that stories were shorter, contained explicit definitions, and drew attention to the topic of word learning. A strong impact of incidental vs. directed learning on lexicalisation is nonetheless unlikely given that such a manipulation (reading-only vs. reading-plus) did not affect the emergence of lexical competition in Experiment 1. To clarify the underlying cause of the discrepancy, a direct comparison of 12 vs. 18 exposures should be made using the same materials. In general, the difference in the strength of lexical competition effects between Experiment 1 and 2 suggests that the prime-lexicality effect could be used as a continuous measure of lexicalisation, rather than merely indicating the presence of absence of lexicalisation (see Chapter 6 for further discussion).

In summary, measures of lexical configuration and lexical engagement of the novel word forms in the present study exhibited different time courses, with lexical configuration emerging after fewer exposures than lexical engagement. This is in line with theoretical distinction between lexical configuration and lexical engagement as two different types of cognitive processes involved in word learning, with different time-courses (Leach & Samuel, 2007). This finding is also consistent with the Complementary Learning Systems framework (McClelland et al., 1995), proposing a fast-learning hippocampal system encoding episodic memories, and a slower consolidation process of these memories involving integration with existing knowledge. While recall and recognition of novel word forms could be achieved on the basis of episodic memories, lexical competition is considered to be dependent on integration with existing word knowledge.

3.4.3 Limitations and conclusions

The design of the current experiment left open the question as to whether the emergence of lexical engagement of written words takes time, or requires a high number of exposures, as the two factors were confounded. Children in the twelve-exposure condition not only received more exposures per novel word, their exposure phase also involved multiple sessions extended over a longer period of time. Lexical configuration as indexed by good knowledge of the word forms was already present after four exposures, yet lexical engagement was not yet evident. The question is which factors were crucial for this knowledge to become integrated with existing knowledge, resulting in the lexical engagement seen after 12 exposures over three sessions. Further exposures in sessions 2 and 3 presumably led to the formation of additional episodic memory traces, while the extension of the exposure phase over a longer period of time could have allowed for additional consolidation of the content of these traces. Further research is necessary to tease apart the role of additional sessions vs. additional exposures. This could be achieved by keeping the number of exposures constant while varying the number of sessions across which they are distributed. If a prolonged exposure phase is crucial for lexicalisation, four exposures spread across multiple sessions should yield evidence of more lexical engagement compared to the four exposures in a single session used in Experiment 2. This would be in line with research showing that distributed practice of novel words can be more effective than massed practice, at least in spoken language. For example, Sobel et al. (2011) exposed 10-year-olds to novel vocabulary in two sessions that were separated either by one minute (massed condition) or one week (spaced condition). Five weeks later children were able to define more words encountered in the spaced condition than in the massed condition. Further research is necessary to investigate whether the same applies to orthographic learning. If however the increase in absolute number of exposures is what allowed lexical engagement in the twelve-exposures condition, then four exposures should not give rise to lexical engagement, regardless of whether they occur in a single session as in Experiment 2, or are distributed over multiple sessions.

Beyond the influence of numbers of exposure, it is also possible that there is a relationship between the relatively weak semantic knowledge of the novel words and the lack of a lexical competition effect in the four-exposures group. Qiao and Forster (2013) only found evidence of lexical competition following a training paradigm that provided participants with definitions of the novel words. While good orthographic and semantic representations have been argued to contribute to the overall lexical quality of a word (Perfetti & Hart, 2002), the role of semantic knowledge in the emergence of lexicalisation of is largely unknown. This issue is discussed in greater detail in Chapter 4, where the provision of semantic information during exposure was manipulated experimentally.

In conclusion, Experiment 2 replicated the prime-lexicality effect in a new sample of children and using improved materials closer to real-life incidental learning, validating the use of this approach for investigating orthographic learning. The present findings are consistent with existing theories suggesting that the integration of novel word knowledge into the lexicon takes longer than the creation of episodic word memories. Further research is necessary to clarify precisely which factors are crucial for the emergence of lexical engagement of novel word forms, including for example a sufficient number of exposures, sufficient time for consolidation, or the support of semantic knowledge.

Chapter 4:

Investigating the Role(s) of Semantics in Orthographic Learning

4.1 Introduction

According to the self-teaching hypothesis, the primary mechanism for orthographic learning is phonological decoding (Share, 1995; 1999). However, research has shown that the relationship between successful decoding of a novel word and subsequent orthographic learning success does not hold strongly at an item level (Nation et al., 2007; Wang et al., 2013). This opens up the possibility that other factors influence the orthographic learning process over and above decoding. There is good evidence that knowledge of meaning affects word reading (for a recent review, see Taylor, Duff, Woollams, Monaghan, & Ricketts, 2015), but less is known about semantic effects in orthographic learning.

A role for semantics is implied in the classic self-teaching paradigm, where target words are embedded in meaningful contexts, but this possibility was not tested directly in the original experiments (Share, 1999). In Experiments 1 and 2 of this thesis, children encountered novel written words that were not previously in their oral vocabulary. Semantic information about the words was provided by story contexts. In Experiment 2, evidence of successful lexicalisation was only evident in the twelve-exposures condition, but not the four-exposure condition. This coincided with poor semantic learning in the four-exposure condition. It is possible that there was a causal relationship between poor semantic learning and lack of lexicalisation in the four-exposures condition. Alternatively, number of exposures could have affected both semantic learning and lexicalisation independently. To test the possibility of a causal relationship, the experiments reported in this chapter manipulated the likelihood of semantic learning directly. Experiments 3.1 and 3.2 compared incidental learning of novel written words in the presence vs. absence of semantic information. As in Experiments 1 and 2, two types of measures of orthographic learning were contrasted: lexical configuration of novel word forms as measured by orthographic decision and spelling, and lexical engagement as indexed by the prime-lexicality effect.

4.1.1 Learning to read words unknown in oral vocabulary

Broadly speaking there are two different scenarios for semantics to influence the learning and lexicalisation of novel word forms. The experiments in this chapter focus on learning of novel written words that are not in children's existing oral vocabularies. A different scenario is given when children encounter written words they have not seen before, but that are familiar to them in spoken language. Several studies have investigated the influence of pre-existing semantic knowledge of written words on learning of their spellings, and the role of context in orthographic learning of familiar words (e.g. Landi et al., 2006; Martin-Chang et al., 2007; McKague, Pratt, & Johnston, 2001; Wang et al., 2011, 2013). An in-depth discussion of this area of research is beyond the scope of this thesis.

Experiments 1 and 2 investigated incidental learning of novel word forms unlikely to be known by children prior to exposure. The same applies to children encountering nonwords in the classic self-teaching paradigm (e.g. Share, 1999). In this case, the initial decoding process cannot be supported by phonological or semantic knowledge of the novel words. However, semantic information provided by context could nonetheless affect orthographic learning. If novel words are embedded in meaningful context, it does become possible to instantiate meaning when first encountering them. In this case, orthographic and semantic (as well as phonological) knowledge of the novel words could be acquired at the same time, as evident in children's word learning in Experiments 1 and 2. In the absence of context or semantic information, e.g. when children read novel words in isolation, only phonological and orthographic representations can be formed. For the present discussion, the key question is whether concurrent semantic learning of novel words impacts orthographic learning and the lexicalisation process. It is possible to imagine both a positive and a negative impact. For example, concurrent semantic learning could limit resources available for orthographic learning (cf. Martin-Chang & Levesque, 2013). However, one could also imagine a mechanism where it is more useful or interesting to learn, or pay attention to, the written form of words that appear to be meaningful, rather than meaningless letter strings. Furthermore, high-quality semantic representations have been argued to contribute to the overall lexical quality of a word (Perfetti & Hart, 2002).

Nation et al. (2007) manipulated the availability of semantic information provided by context in a self-teaching paradigm. Children aged 7-8 years read aloud monosyllabic nonwords either embedded in stories or in isolation. There was no effect of context as measured by orthographic choice, both after a one-day delay and after seven days. It is important to note the study used monosyllabic nonwords assigned with a regular pronunciation, so they might not generalise to words with different orthographic or phonological properties.[5] It is also possible that the parameters of the context matter – for instance, the ease with which a specific meaning can be inferred or the semantic properties of the word, e. g. the richness of the semantic information provided.

---

[5] In order to manipulate the regularity of novel words in a self-teaching paradigm, children have to pre-exposed to the correct pronunciations – otherwise they are likely to regularise all novel words (see Wang et al., 2011, 2013). This is then akin to orthographic learning of words known in oral vocabulary, which is different from learning of novel words as studied in the present experiments.

A beneficial effect of context was demonstrated in adults by Bolger et al. (2008). Participants encountered rare words, some in sentence contexts and some in isolation. A benefit of semantics was supported by the observation of better orthographic choice performance for words encountered in context than in isolation. This evidence was further strengthened by an item-level relationship between participants' ability to generate word meanings at post-test and orthographic choice accuracy. Apart from potential differences between adults and children, there were further differences in the training paradigms between Bolger et al. (2008) and Nation et al. (2007), which could explain the discrepant results. While Nation et al. (2007) used an incidental learning paradigm, Bolger et al. (2008) instructed participants to 'do their best to learn the meanings of the words from context as they would be assessed later' (p. 138). Thus, participants were more likely to draw inferences about the word meanings, and generally they may have paid more attention to words seen in context than those in isolation, where no meaning was available. Furthermore, words were embedded in sentences rather than stories. While all semantic information provided by Bolger et al.'s (2008) sentence contexts related to the novel words, this was not the case for Nation et al.'s (2007) stories. Further research is necessary to clarify whether these differences could explain why a semantic benefit for orthographic learning was seen by Bolger et al. (2008) but not in Nation et al.'s (2007) study.

Ouellette and Fraser (2009) provided further evidence consistent with the idea that the salience or accessibility of semantic information provided for novel words could be of importance. In their experiment, novel semantic information was provided for nonwords when they were first encountered in written form, without the need for children to infer meaning from context. Children aged 9-10 years were exposed to ten monosyllabic nonwords; five in a semantic condition and five in a non-semantic condition. In the semantic condition, children were shown each word written on a card, asked to repeat it after the experimenter, listened to its definition and were shown a matching picture. Children were then asked to read aloud each word a further six times and point to the corresponding picture. Hence, semantic information was provided more explicitly than in Nation et al.'s (2007) study, where children had to infer word meanings from context. Nonetheless, semantic information was provided concurrently while children were exposed to novel word forms, rather than established in children's oral vocabulary beforehand (cf. Wang et al., 2011). In the non-semantic condition, children read aloud the nonwords and made lexical decisions, receiving the same number of exposures per nonword as in the semantic condition. Children were tested on orthographic choice and spelling one day later and four days later. At both time points, there was a small-to-moderate effect of exposure condition on orthographic choice. Children performed better on nonwords they encountered in the semantic condition than in the non-semantic condition. There was no difference between exposure conditions on spelling performance. This provides some evidence that providing semantic information as children encounter a novel word in writing aids orthographic learning. However, it is questionable how well this paradigm reflects children's experiences outside the laboratory. Very early readers might encounter picture books where pictures are labelled with individual words, or they might attempt to read words used as labels on products, buildings, etc., which bears some resemblance to Ouellette and Fraser's (2009) paradigm. However, it seems likely that as soon as children are able to read connected text, the availability of semantic information about novel written words will be dependent on the nature of written contexts rather than pictures or spoken definitions. Hence, in the present experiment, semantic information was provided in an incidental learning paradigm, where novel word meanings have to be inferred from context, as in

Nation et al. (2007). However, in contrast to that study, the present experiment used rare words rather than nonwords, and made use of longer story contexts providing richer semantic information for the novel words, albeit in an incidental and non-directed fashion (as in Experiments 1 and 2 of this thesis).

In summary, for written words that are previously unknown, so far, semantic information provided simultaneously has been shown to enhance orthographic learning in children only when it is explicit and easily accessible. Hence, it is possible that the time course of semantic learning is a moderator of its effect on orthographic learning. In Ouellette and Fraser's (2009) study, children were able to acquire the nonword meanings very quickly and were able to identify the corresponding pictures without errors after two of six exposures to each nonword. Hence the semantic representations formed during the first two exposures may have assisted children's decoding efforts during subsequent exposures. In contrast to this, Experiment 2 (Chapter 3) revealed that learning word meanings incidentally from contexts is comparatively slower. Here, children were able to define only a few novel words after a single session with four exposures per word. Semantic learning was significantly better if children encountered novel words over three sessions with twelve exposures per word. This could be explained by the need to draw inferences in order to learn about novel word meanings from context. Only if and when such inferences are drawn and semantic information is successfully acquired can it aid subsequent word reading processes. It is possible that in Nation et al. (2007), semantic learning was either too weak or too slow for it to impact orthographic learning within a single session with a maximum of four exposures per word.

In the present experiment, the exposure phase involved three sessions totaling twelve exposures per word embedded in rich story contexts. Here, the clean-cut case of words being novel in both form and meaning is probably only given at the very first exposure. During all subsequent exposures, newly-formed semantic representations of the word might begin to influence the word recognition/ decoding process (akin to experiments investigating oral pre-exposure and in particular for irregular words, e.g. Wang et al., 2011) and by extension serve to influence further strengthening of orthographic representations and lexicalisation processes. Semantic representations could also evolve, e.g. be strengthened by re-activation, or extended if a novel aspect or nuance is introduced (the role of variability in semantic information for novel words is explored further in Chapter 5). Thus it becomes difficult to tease apart effects of existing semantic knowledge vs. effects of on-going semantic learning, and semantic influences on orthographic learning might change over the course of multiple exposures.

4.1.2 Semantics in masked priming

Effects of semantics might also be influenced by whether or not the activation of semantic representations could influence performance on the chosen measure of learning or lexicalisation itself. A measure involving explicit word recognition such as reading aloud or orthographic choice allows for semantic representations to be activated (if they exist), so performance might not solely reflect the quality of orthographic representations. The test itself confounds the influence of semantics on activation of orthographic representations and on how well they were established. Hypothetically, what might look like superior orthographic *learning* might actually constitute superior *identification* at test, due to non-orthographic factors such as semantic influences on activation.

One way to circumvent this problem is to use a measure that minimises the activation of semantic representations, resulting in a ‘purer’ measure of orthographic representations. This is argued to be the case in masked priming, as the brief and masked presentation of the prime eliminates on-line activation of its semantic representations (reviewed in Kouider & Dehaene, 2007). Hence, the prime-lexicality effect was used in the present experiments for two reasons. First, the experiments explored the influence of semantic factors on the lexicalisation of novel words as indexed by the prime-lexicality effect in comparison to their lexical configuration (which was also assessed). This comparison has not been made before, and the effect of semantic factors on lexical engagement of novel written words is yet unknown. Second, the prime-lexicality effect offers an insight into orthographic representations that is not contaminated by concurrent activation of semantic representations.

So far, the availability of semantic information and its influence on orthographic learning has not yet been manipulated with regard to the prime-lexicality effect. Qiao and Forster (2013) included semantic information of the novel words in their training paradigm with adults and found evidence of novel word lexicalisation after four training sessions, but this was not compared to a non-semantic condition. Hence, it is not clear whether lexical engagement requires extended training, or meaning, or indeed both. Experiment 2 (Chapter 3) revealed different time courses of lexical configuration of forms, lexical configuration of meaning, and lexical engagement. While lexical configuration of word forms was already apparent after four exposures, good semantic knowledge was not yet evident. Semantic knowledge was better after twelve exposures, when lexical engagement also emerged. Based on these findings, one could speculate that semantic information is more important for lexicalisation as indexed by lexical engagement than lexical configuration. This would be in line with the lexical quality hypothesis (Perfetti & Hart, 2002), which emphasises the importance of high-quality semantic, orthographic and phonological representation for each word in the lexicon. The present experiments are the first to investigate whether the availability of semantic information influences the learning of novel word forms both as measured in the self-teaching tradition, and on lexical engagement as measured by the prime-lexicality effect. As in previous experiments, the exposure phase was extended over several sessions and exposures, allowing for potential dynamic interplay between evolving semantic and orthographic representations.

4.2 Experiment 3.1

Experiment 3.1 compared two conditions. In the semantic condition, novel words were embedded in story contexts read aloud by the children, as in Experiment 2. Based on pilot data (see Chapter 2), children were assumed not to have the novel words in their oral vocabulary prior to the experiment. In the non-semantic condition, novel words were read aloud in meaningless lists of words. Here it was assumed that the ease of reading could not be influenced by prior semantic representations, as words were likely unknown by the children, nor could any novel semantic learning take place, as no semantic information was provided.

Hypotheses

Given the novelty of the paradigm and the mixed findings in the literature, a number of different outcomes could be hypothesised. No firm predictions were made to differentiate the effects on lexical configuration and lexical engagement, so in the following ‘orthographic learning’ is used to refer to both.

A) Since Experiment 2 showed some semantic learning after one exposure session, it was possible that newly created semantic representations could begin to influence the reading process in the semantic condition after the first few exposures. Evidence for this would be higher reading accuracy on the novel words in the semantic condition compared to the non-semantic condition, in particular later in training. In this case, the increased reading success in the semantic condition could result in better orthographic learning compared to the non-semantic condition.

B) Alternatively, orthographic learning might be equal for both conditions as exposure to orthography and/or lexical phonology (provided in both conditions) could be sufficient for orthographic learning. This outcome could also arise if the benefits of semantics were cancelled out by a confounding factor, such as an attentional advantage of processing words presented in isolation, or greater task congruency between exposure and test for the non-semantic condition (single word presentation at exposure and test).

C) Third, orthographic learning could be superior in the non-semantic condition. This could be due to a negative effect on available attentional resources by processing of the context and/or parallel semantic learning in the semantic condition, or lack attention to individual letters due to the increasing predictability of the words from context, as the same stories were read repeatedly.

4.2.1 Method

4.2.1.1 Design

The experiment involved a mixed design with a between-subjects manipulation of exposure condition (semantic vs. non-semantic). The experiment consisted of three phases – pre-exposure, exposure phase, and post-exposure. Data for masked priming lexical decision were collected before and after, yielding a further within-subject factor of test session (pre-exposure vs. post-exposure). Reading accuracy of the novel words was recorded for the first exposure to each word in each of the three exposure sessions. Spelling, orthographic decision, and definition knowledge data were collected once, after exposure.

4.2.1.2 Participants

Twenty children (11 female, 9 male) from Year 5 and Year 6 classes participated in the experiment. They were aged between 9.2 and 11.7 years (M = 10.58 years, SD = .74 years). They were selected on the basis of being monolingual native English speakers scoring no more than 1SD below the standardised average on the Test of Word Reading Efficiency (TOWRE; Torgesen et al., 1999). This test consists of two subtests – word reading and nonword reading. Performance on the two subtests was highly correlated (r = .80). The sample consisted of 10 children in the semantic group and 10 children in the non-semantic group[6]. Descriptive statistics are shown in Table 4.1. The

[6] For the semantic group, data from the 10 children collected as part of Experiment 4 was used. For the non-semantic group, data from 10 additional children was collected.

overall mean total standard TOWRE score was 109.72 (*SD* = 17.77), indicating that children performed within the population average. Groups did not differ significantly in age or age-standardised reading ability as indexed by the TOWRE. They did however differ in raw total TOWRE scores (see Table 4.1). Further examination of the TOWRE subtests revealed that the two groups differed significantly on nonword reading ability ($p = .018$), but not word reading ability ($p = .122$). All analyses controlled for this difference by including reading ability as a covariate. Children were recruited using an opt-out procedure, where parents were sent two information sheets over the course of two weeks with the opportunity to opt out of their child's participation in the study by returning a response slip.

Table 4.1. Descriptive characteristics of semantic and non-semantic groups in Experiment 3.1

	Semantic (N = 10)		Non-semantic (N = 10)		Group comparison
Measures	*M*	*SD*	*M*	*SD*	*p*
Age (years)	10.68	.83	10.49	.77	.595
TOWRE total raw score[a]	117.5	17.19	101.5	15.14	.040
TOWRE total standard score[b]	110.9	11.74	102.3	12.98	.137

Notes. [a] total number of words read correctly on both subtests in 45 seconds per test, [b]$M = 100$, $SD = 15$.

4.2.1.3 Materials and procedure

Word stimuli

The 20 word pairs were identical to those used in Experiments 1 and 2, comprising novel words and their orthographic neighbours (base words). For counterbalancing purposes, words were once again divided into lists A and B.

Pre-exposure baseline assessments

Masked priming lexical decision. This task was identical to that used in Experiment 2. Due to scheduling constraints, some children completed this task on the same day as the start of the exposure phase, whereas other completed the task on a separate day before the exposure phase.

Exposure phase

The exposure phases consisted of three sessions completed on separate days. The number of days across which the exposure phase was completed varied from 3 to 16 days (M = 6.75, SD = 4.15). This did not differ between the two groups (p = .959).

Materials. For the semantic condition, the same story contexts were used as in Experiment 2. For the non-semantic condition, word lists were created on the basis of these stories. Like the stories, each list contained one of the novel words four times. Words from the stories were listed in random order in two columns on A5 pages. Some words, such as function words and names, were excluded to arrive at a mean list length of 61 words. Repetitions of filler words were reduced to a maximum of four (for full lists, see Appendix 4.1).

Exposure conditions.

Semantic condition. The semantic condition was identical to the twelve-exposures condition of Experiment 2, involving three sessions spread over different days with four exposures per novel word per session (12 exposures in total).

Non-semantic condition. The non-semantic condition also involved three sessions with four exposures per novel word per session (12 exposures in total). On each day, children were asked to read aloud lists of words. They were told that these might contain some unfamiliar words. Children were provided with the correct pronunciations of the target words if they read them incorrectly. To keep children engaged with the task, filler games were created. In session 1, children completed a puzzle and were allowed to pick and place one piece after each list. In session 2, children deciphered a secret message that was gradually revealed after each list. In session 3, children completed a variation of the 'Cluedo' game, where one suspect was eliminated after each list. They were also always given a longer break after half of the lists, when they engaged in a short non-verbal task, such as origami or mazes.

Assessment of reading accuracy during exposure. On each of the three exposure days, children's first attempt at reading each of the novel words was scored as either correct or incorrect. Thus, for each day there was one accuracy score for each of the ten novel words per child.

Post-exposure assessments

The procedure at post-test was identical to Experiment 2. Children were tested on definition knowledge, masked priming lexical decision, orthographic decision, and spelling**.** Children were tested between 1 and 5 days after completing the exposure phase (*M* = 1.6, *SD* = 1.05), depending on their availability. The test delay did not differ between the two groups ($p = .090$).

4.2.2 Results

Analyses were performed using linear mixed effects models following the same principles outlined in Chapter 2. For models with continuous outcome variables, significance was determined using the cut-off point of $|t| > 2$, based on recommendations by Barr et al. (2012). For models with binary outcome variables, mixed effects logistic regression models were computed. Here, the cut-off point for significance was $p < .05$. To control for individual variation and the group difference in word reading ability, children's total raw TOWRE scores were included as covariates in the analyses.

4.2.2.1 Novel word reading accuracy during exposure

Reading accuracy scores were recorded for the first exposure of each novel in each exposure session. Mean scores are shown in Figure 4.1. To compare children's reading accuracy for the novel words across the two exposure conditions, a generalised logistic mixed model was fit, with exposure condition (semantic vs. non-semantic) as a fixed factor. Models including session as a fixed factor did not converge, so this was not included in the final model (see Appendix 4.2). Numerically, performance increased across exposure sessions, as shown in Figure 4.1. The main effect of exposure condition ($p = .064$), indicated a trend towards higher accuracy in the semantic condition than in the non-semantic condition. TOWRE score was not a significant covariate ($p = .078$).

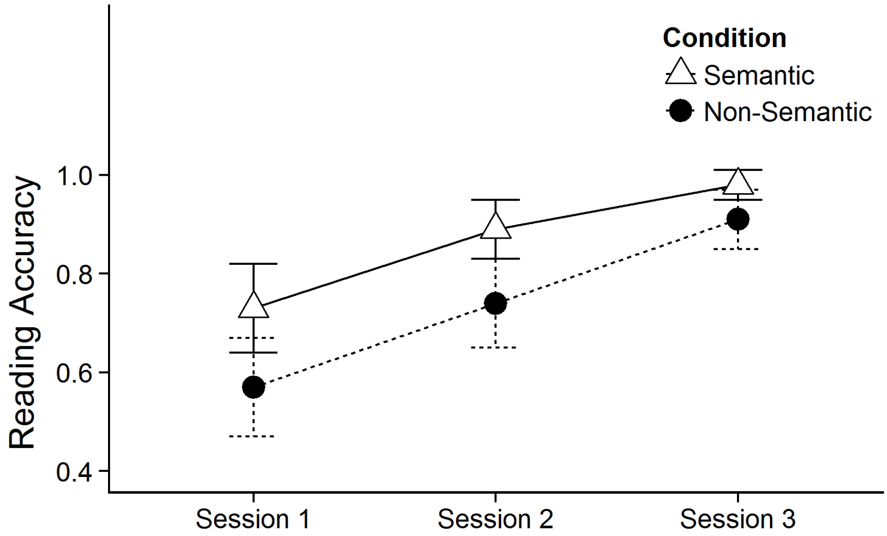


Figure 4.1. Mean reading accuracy scores across exposure sessions per exposure condition. Error bars show 95% confidence intervals.

4.2.2.2 How well did children learn the novel words?

Three tasks assessed how well the novel words were learned. Definition knowledge assessed semantic learning and served as a manipulation check of the distinction between semantic vs. non-semantic exposure conditions. Orthographic decision and spelling assessed form learning. Figure 4.2 shows mean accuracy per exposure condition for each task, indicating high accuracy levels across tasks, with the exception of definition knowledge in the non-semantic condition, which was at floor.

To compare performance between exposure conditions, generalised logistic mixed effects models were fit separately for each task. Exposure condition (semantic vs. non-semantic) was entered as fixed factor. For definition knowledge, there was a significant main effect of exposure condition. Children in the semantic condition could define a greater proportion of words correctly than those in the non-semantic condition (p <.001). TOWRE score was not a significant covariate (p = .538). This confirmed that the semantic vs. non-semantic exposure manipulation was successful. There was no difference between exposure conditions on either of the two measures of form learning. Both groups performed equally on orthographic decision (p = .413). TOWRE score was a significant covariate (p = .007), indicating that children with better reading skills scores performed more accurately on the task. There was also no difference between exposure conditions in the ability to spell words correctly (p = .545); here, TOWRE score was not a significant covariate (p = .362). In summary, there was a difference between exposure conditions was for definition knowledge, but not orthographic decision or spelling.

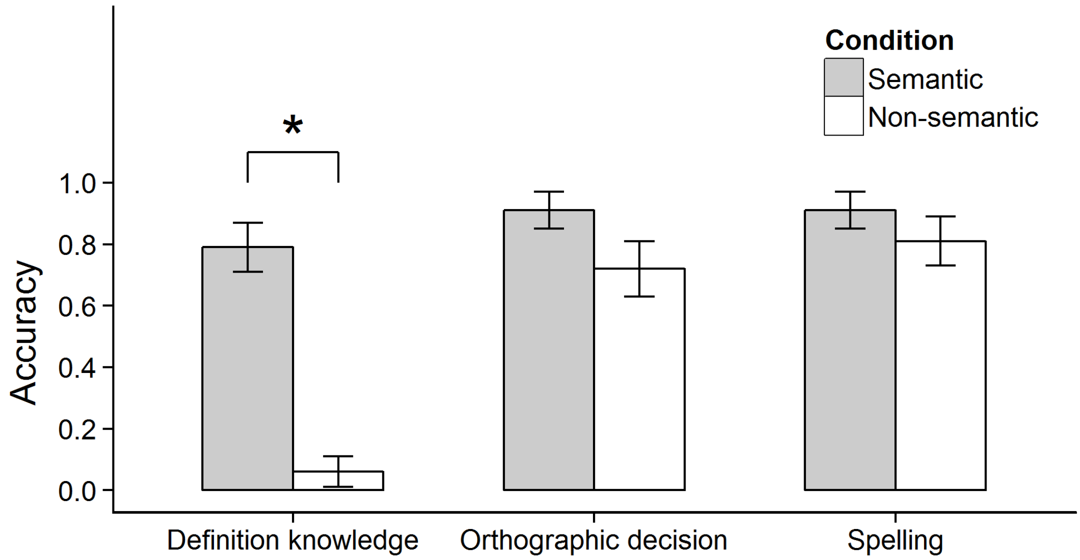


Figure 4.2. Mean performance on definition knowledge, orthographic decision, and spelling per exposure condition. Error bars show 95% confidence intervals.

4.2.2.3 Did the novel words induce lexical competition effects on processing of their neighbours?

Masked priming lexical decision. Due to an error in the experimental script, data from 40 trials (5%) had to be excluded from the analysis. Only word trials were included in the analysis. Overall, children performed with high accuracy on this task, with errors made on 53 out of 760 (6.97%) trials across sessions. This was deemed insufficient for further analysis. Based on visual inspection of the qq-plot, RT values were inverse transformed to reduce positive skew, and one data point was removed. Figure 4.3 summaries the data per condition. For ease of interpretation, raw RT data are plotted. Inverse transformed data as used in the analyses are shown in Appendix 4.

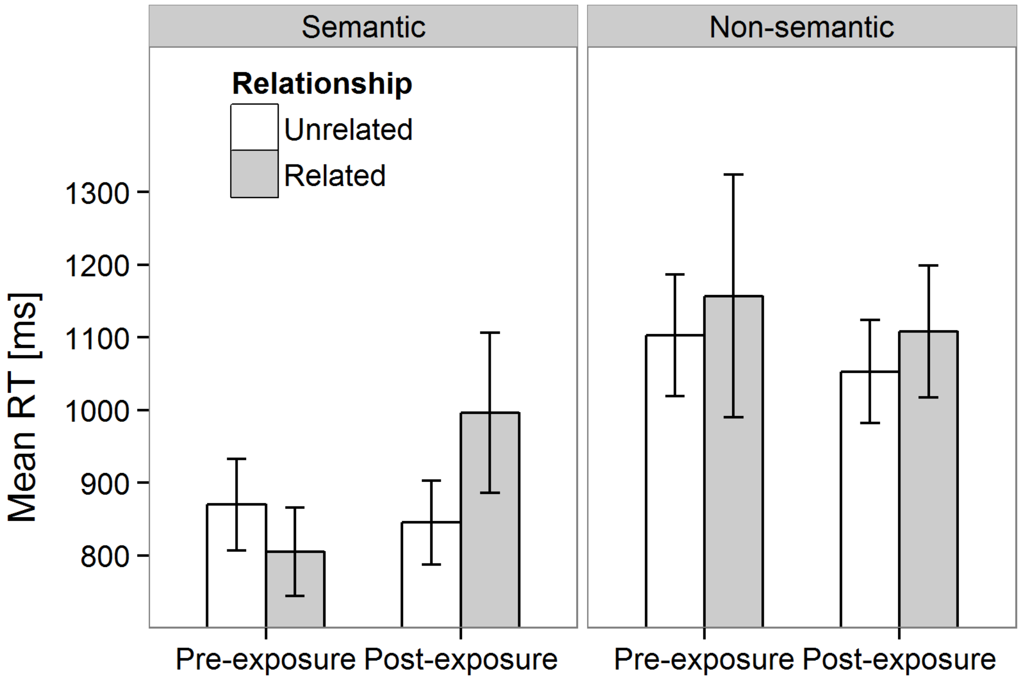


Figure 4.3. Mean RTs on masked priming lexical decision task, split by exposure condition. Error bars show 95% CIs.

To analyse RT, a linear mixed effects models was computed with exposure condition (semantic vs. non-semantic), test session (pre- vs. post-exposure), and prime-target relationship (related vs. unrelated) as fixed factors. TOWRE score was a significant covariate (t = -6.43), showing that children with higher reading ability showed overall faster RTs than children with lower reading ability. The main effects of exposure condition ($t = 0.68$), test session ($t = 0.94$), and prime-target relationship ($t = 0.04$), were not significant. There was however a significant prime-target relationship x test session interaction ($t = 2.91$, Figure 4.4). There were no further significant interactions (see Appendix 4.2 for full models).

To explore the prime-target relationship x test session interaction, separate models were computed for pre-exposure and post-exposure data, with prime-target relationship as a fixed factor. This revealed a trend towards faster responses to targets primed by related novel words than unrelated words prior to exposure ($t = -1.93$). After exposure, the effect of prime-target relationship was not significant ($t = 1.37$). TOWRE was a significant covariate both at pre-exposure ($t = -6.44$) and post-exposure ($t = 4.42$), indicating that children with higher TOWRE scores showed overall faster RTs than children with lower scores. In summary, the interaction between test session and prime-target relationship indicated a switch in direction of priming effects from nonword-like to word-like priming behaviour. This was not qualified by exposure condition.

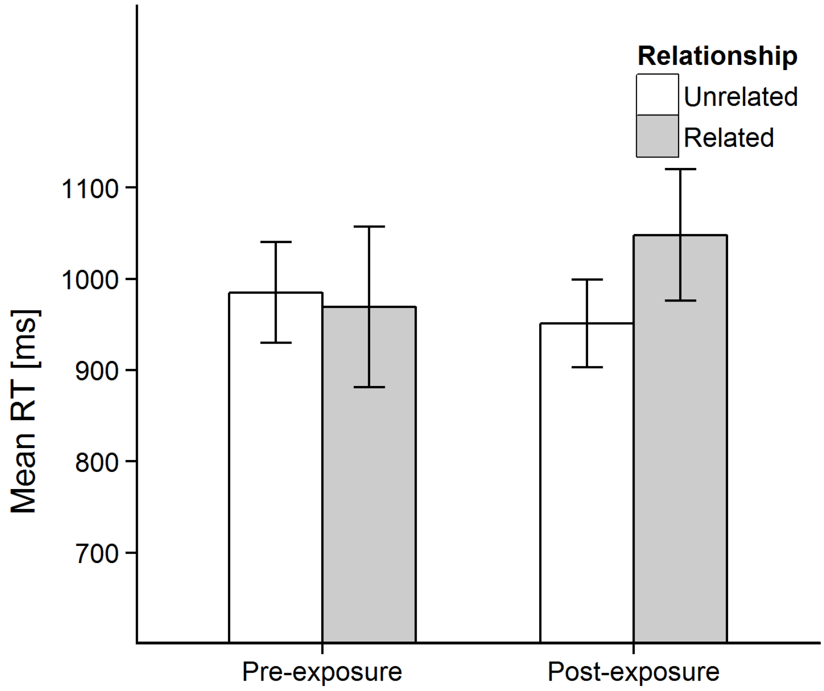


Figure 4.4. Mean RTs for the masked priming lexical decision task collapsed across exposure conditions, showing the interaction between prime-target relationship and test session. Error bars show 95% CIs.

4.2.3 Interim Discussion

Experiment 3.1 revealed no difference in form learning and lexicalisation between a semantic and a non-semantic condition. This suggests that semantics provided by context are not needed for these processes, contrary to what the findings of Experiment 2, and Qiao and Forster (2013), might suggest. The question remains what factors allowed for these good levels of learning and lexicalisation. While the non-semantic condition did not provide children with semantic knowledge of the novel words, several factors might have enhanced the lexical phonology and hence lexical status of the novel words. First, it is possible that the provision of feedback on children's pronunciation of the novel words enhanced their lexical status in the non-semantic condition, making the contrast with the semantic condition less stark. Second, the definition knowledge task was administered before the masked priming lexical decision task at post-exposure test. This not only provided an exposure to the orthographic forms of the novel words, it also implied that the novel words were indeed words with meanings, rather than nonsense words. These factors might have driven learning and lexicalisation to be equal to the semantic condition. Finally, children in the two conditions were not perfectly matched on overall reading ability, with slightly better readers in the semantic condition, although this was statistically controlled in the analysis.

Experiment 3.2 served as a replication of Experiment 3.1 in a new sample of children, while addressing the factors outlined above. Children did not receive feedback on their pronunciations and were tested on masked priming lexical decision as the first task at post-exposure. If the provision of feedback on the phonological forms of the novel words was indeed critical for lexicalisation in the non-semantic condition, the prime-lexicality effect should only emerge for the semantic condition in the new experiment. If however reading aloud novel words embedded in word lists without feedback allows equal lexicalisation as reading novel words in story contexts, a prime-lexicality effect should emerge across exposure conditions, as in Experiment 3.1.

4.3 Experiment 3.2

4.3.1 Method

4.3.1.1 Design

The design was identical to Experiment 3.1, involving a between-subjects manipulation of exposure condition (semantic vs. non-semantic).

4.3.1.2 Participants

Twenty-one children (9 female, 12 male) from a Year 6 class participated in the experiment. They were aged between 10.23 and 11 years (*M* = 10.59 years, *SD* = .28 years). As in Experiment 3.1, they were selected on the basis of being monolingual native English speakers scoring no more than 1*SD* below the standardised average on the Test of Word Reading Efficiency (TOWRE; Torgesen et al., 1999). Performance on the two subtests was highly correlated ($r = .82$). The sample consisted of 10 children in the semantic group and 11 children in the non-semantic group. Descriptive statistics are shown in Table 4.2. The overall mean total standard score was 114.01 (*SD* = 13.87), indicating that children performed at the upper end of the population average. Groups did not differ significantly in age or reading ability (Table 4.2). The recruitment procedure was identical to Experiment 3.1.

Table 4.2. Descriptive characteristics of semantic and non-semantic groups in Experiment 3.2

	Semantic (N = 10)		Non-semantic (N = 11)		Group comparison
Measures	*M*	*SD*	*M*	*SD*	*p*
Age (years)	10.56	.33	10.61	.25	.666
TOWRE total raw score[a]	121.90	17.14	115.30	14.97	.619
TOWRE total standard score[b]	118.27	15.77	112.91	13.41	.704

Notes. [a] total number of words read correctly on both subtests in 45 seconds per test, [b]*M* = 100, *SD* = 15.

4.3.1.3 Materials and procedure

Pre-exposure baseline assessments

Masked priming lexical decision. This task was identical to that used in Experiment 3.1.

Exposure phase

The exposure phases consisted of three sessions completed on separate days. The number of days across which the exposure phase was completed varied from 3 to 8 days (M = 3.38, SD = 1.16). This did not differ between the two groups (p = .156). The materials used were identical to Experiment 3.1.

Exposure conditions. As in Experiment 3.1, children were allocated either to the semantic condition or the non-semantic condition. The procedure for both conditions during the exposure sessions was identical to Experiment 3.1, which the exception that children were not given feedback on their reading. Reading accuracy of the novel words was not recorded during the exposure phase.

Post-exposure assessments

The order of post-exposure assessments was different from Experiment 3.1. Children were tested in the order of masked priming lexical decision, spelling, and definition knowledge. Children's reading accuracy of the newly learned words was also recorded during the definition knowledge task. Children were asked to read each novel word from their allocated list printed on a flashcard and then provide a definition. This order of tasks ensured that children did not have an additional exposure to the novel words prior to the masked priming lexical decision task. Data from the orthographic decision task had to be discarded due to experimenter error in how the task was administered. Children were tested between 1 and 3 days after completing the exposure phase (*M* = 1.10, *SD* = .44), depending on their availability. The test delay did not differ between the two groups (*p* = .353).

4.3.2 Results

4.3.2.1 How well did children learn the novel words?

Three tasks assessed how well novel words were learned. Definition knowledge assessed semantic learning and served as a manipulation check of distinction between semantic vs. non-semantic exposure conditions. Reading accuracy at test and spelling assessed form learning. Figure 4.5 shows high accuracy levels across tasks for both exposure conditions, with the exception of definition knowledge in the non-semantic condition, which was approaching floor levels.

To compare performance between exposure conditions, generalised logistic mixed effects models were fit separately for each task. Exposure condition (semantic vs. non-semantic) was entered as a fixed factor. For definition knowledge, there was a significant main effect of exposure condition. Children in the semantic condition could define significantly more words correctly than those in the non-semantic condition (*p* <.001). This confirmed that the semantic vs. non-semantic exposure manipulation was successful. TOWRE score was a significant covariate (*p* = .013), indicating that children with higher TOWRE scores performed better at definition knowledge than children with lower scores. There was no difference between exposure conditions on either of the two measures of form learning. For the analysis of reading accuracy at test, the maximal model failed to converge, so the random effects structure was simplified (see Appendix

4.3). Both groups performed equally on reading accuracy at test (*p* = .951). TOWRE score was a significant covariate (*p* = .024), indicating that children with better general word reading efficiency scores read the novel words more accurately at the post-exposure test. There was also no difference between exposure conditions in the ability to spell the words correctly (*p* = .465). TOWRE score was a significant covariate (*p* < .001), indicating that children with higher TOWRE scores performed better at the spelling task than children with lower scores. In summary, there was a difference between exposure conditions for definition knowledge, but not reading accuracy or spelling.

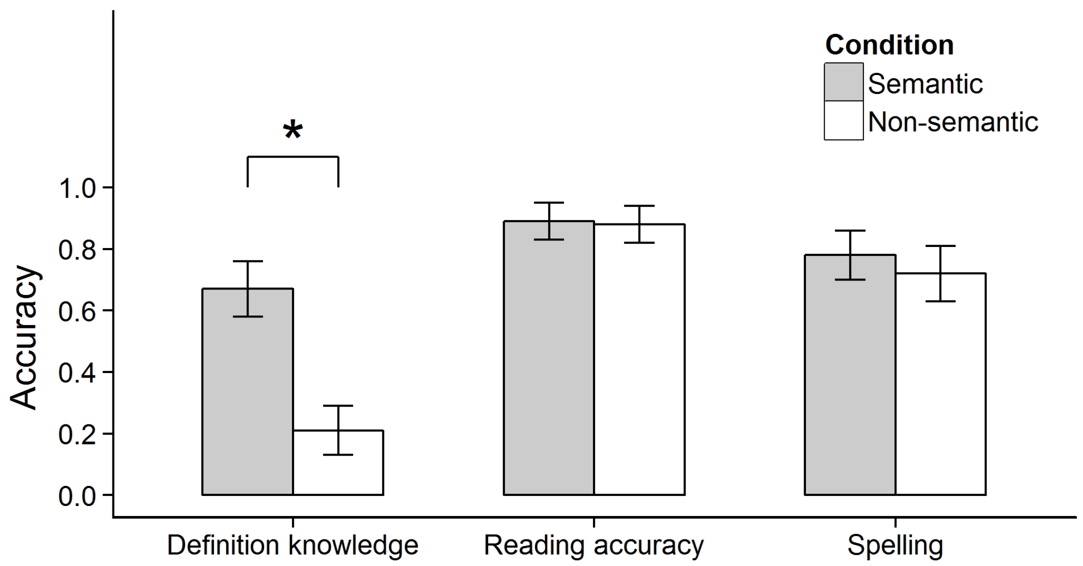


Figure 4.5. Mean performance on definition knowledge, reading accuracy, and spelling per exposure condition. Error bars show 95% confidence intervals.

4.3.2.2 Did the novel words induce lexical competition effects on processing of their neighbours?

Masked priming lexical decision. Overall, children performed with high accuracy, as in previous experiments. Errors were made on 43 out of 840 (5.12%) trials across sessions and this was deemed insufficient for further analysis. Only trials with correct responses were included in the RT analysis. Based on visual inspection of the qq-plot, RT values were inverse transformed to reduce positive skew, and one data point was removed. Figure 4.6 summaries the data per condition. For ease of interpretation, raw RT data are plotted. Inverse transformed data as used in the analyses are shown in Appendix 4.

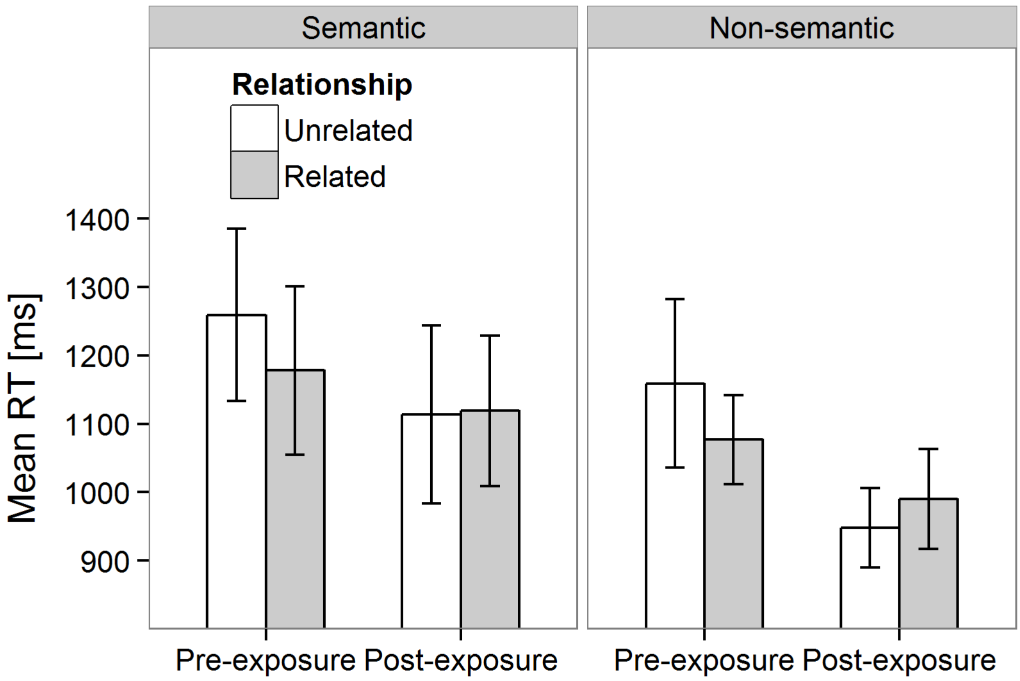


Figure 4.6. Mean RTs on masked priming lexical decision task, split by exposure condition. Error bars show 95% CIs.

To analyse RT, a linear mixed effects models was computed with exposure condition (semantic vs. non-semantic), test session (pre- vs. post-exposure), and prime-target relationship (related vs. unrelated) as fixed factors. TOWRE score was not a significant covariate (t = -1.04). The main effects of exposure condition (t = -1.36) and prime-target relationship (t = -1.15) were not significant. There was however a significant main effect of test session (t = -4.22), indicating overall faster RTs at post-exposure than at pre-exposure. This main effect was qualified by a significant prime-target relationship x test session interaction (t = 2.19, Figure 4.7). There were no further significant interactions (see Appendix 4.3 for full models).

To explore the prime-target relationship x test session interaction, separate models were computed for pre-exposure and post-exposure data, with prime-target relationship as fixed factor. This revealed significantly faster responses to targets primed by related novel words than unrelated words prior to exposure (t = -2.45). TOWRE score was not a significant covariate (t = -1.80). After exposure, the effect of prime-target relationship was not significant (t = 0.66). TOWRE score was not a significant covariate (t = -0.90). In summary, the interaction between test session and prime-target relationship indicated a switch in direction of priming effects from nonword-like to word-like behaviour. This was not qualified by exposure condition.

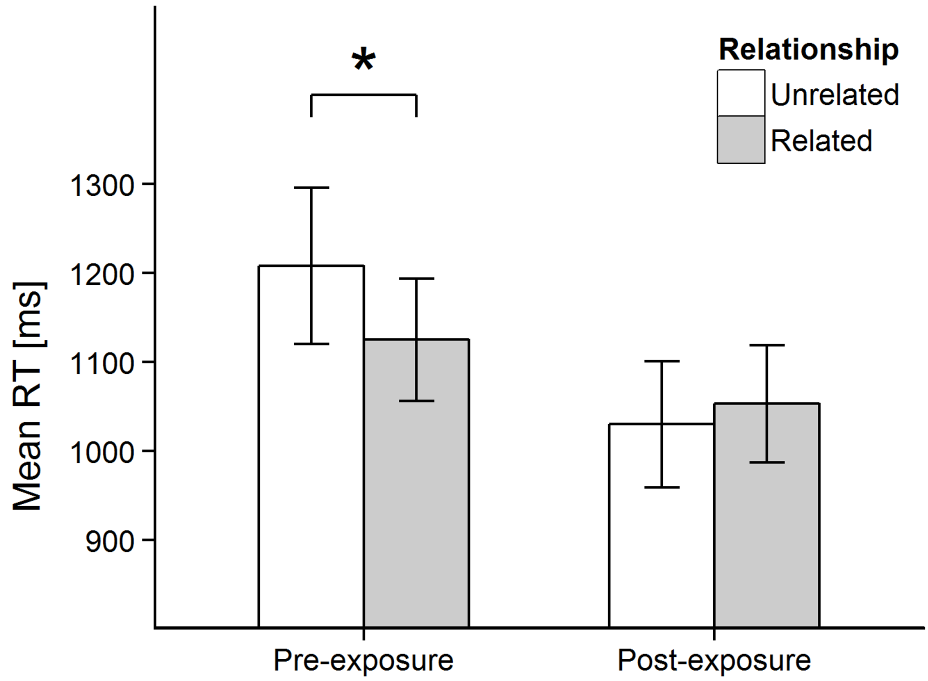


Figure 4.7. Mean RTs for the masked priming lexical decision task collapsed across exposure conditions, showing the interaction between prime-target relationship and test session. Error bars show 95% CIs. Significance is indicated for analyses performed on transformed data.

4.3.3 Interim Discussion

Experiment 3.2 replicated the findings of Experiment 3.1, showing equal orthographic learning and lexicalisation for semantic and non-semantic condition. Importantly, the pattern of findings maintained in the absence of feedback on children's pronunciations of the novel words. This suggests that feedback was not driving the lexicalisation of novel words in the non-semantic condition of Experiment 3.1. Instead, the two experiments revealed robust learning and lexicalisation of novel word forms regardless of whether semantic information was provided by context. This finding is discussed in more detail below.

4.4 Discussion

The present experiments manipulated whether semantic information was provided when children encountered novel written words. The manipulation was effective, evidenced by the fact that semantic learning was significantly higher in the semantic condition than in the in the non-semantic condition in both experiments. In Experiment 3.1, definition knowledge in the non-semantic learning approached floor levels. In the non-semantic condition, five children out of ten were able to define one of the words; one child was able to define two. It is likely that these children were already familiar with those words prior to the experiment. Children confirmed this anecdotally. In Experiment 3.2, definition knowledge in the non-semantic condition was numerically higher than in Experiment 3.1 (M = .06 vs. M = .21), but still significantly lower than in the semantic condition. Again, it is likely that children's ability to define some of the words was due to prior knowledge rather than semantic information provided by the word lists, despite the careful piloting of words reported in Chapter 2. This illustrates the difficulty of choosing a set of words that will be novel to a large range of children[7].

A consistent finding emerged from the present experiments, both revealing equal levels of orthographic learning across semantic and non-semantic conditions. This was evident both for measures of lexical configuration (orthographic decision and spelling) and lexical engagement as measured by the prime-lexicality effect. Second, in Experiment 3.1, during the exposure phase, children in the non-semantic condition showed a trend towards lower reading accuracy of the novel words than children in the

[7] Assessing knowledge of the novel words for the children in the sample prior to the experiment would offer them an additional, non-incidental exposure and could potentially highlight word-learning as the objective of the experiment.

semantic condition, controlling for general word reading efficiency, a finding discussed in the following.

4.4.1 Effects of emerging semantic knowledge on reading accuracy

Previous research has shown that context aids reading accuracy for words familiar in oral language compared to when words are read in isolation (Landi et al., 2006; Martin-Chang et al., 2007). This might be because context activates existing semantic and phonological representations, which could aid the decoding process. These findings are broadly compatible with Experiment 3.1, where a trend towards a benefit of context over isolation for new word reading emerged, collapsed across exposure sessions. Numerically, reading accuracy appeared higher for reading in context than reading in lists from the very first exposure session. Arguably, at this point children in both conditions should have neither lexical phonology nor semantic representations for the words. The difference might have been due to general differences in reading ability between the two groups, which might not have been sufficiently captured by controlling for TOWRE scores as a measure of word reading fluency. A benefit of reading words in context would be expected to emerge only in later exposure sessions, when novel word meanings became familiar, and could in turn be activated by the context to aid decoding. Further research using a larger sample is necessary to pinpoint the emergence of a benefit of reading in context compared to isolation for novel words, and to investigate whether this benefit can indeed be linked to emerging semantic learning of the words.

4.4.2 Effects of semantics on knowledge of novel word forms and their lexicalisation

In both experiments, the provision of novel semantic information in story contexts did not benefit orthographic learning compared to encountering just the written forms. This was true both regarding lexical configuration, i.e. children's knowledge of spellings and their ability to read words correctly at post-test, as well as lexicalisation of word forms as indexed by the prime-lexicality effect. There are several further possible explanations for this finding.

As outlined in the introduction to this chapter, when exposing children to novel words over time, it is difficult to tease apart effects of on-line semantic learning early in the exposure phase and activation of newly established semantic information later in the exposure phase. Children in the semantic condition may have been at a disadvantage due to the processing demands of on-line semantic learning early in the exposure phase, which could have cancelled out any benefit of semantic support potentially emerging later in the exposure phase.

More generally, the difference in the processing demands of reading connected, meaningful texts, and reading individual words in meaningless lists might constitute a potential confound. It has previously been argued that the provision of context influences the way children decode words, such that context could aid the word reading process to the point that it becomes unnecessary to pay attention to the precise sequence of letters (Landi et al., 2006). Arguably, this has only been shown for words likely to be familiar in oral vocabulary, prior to their encounter in print. This is quite different from the present study, where at least initially, context provided semantic information about the novel words, rather than allowed activation of existing semantic knowledge. More likely, the difference in format (context vs. isolation) could have introduced effects of task transfer, as all measures of orthographic learning and lexicalisation were based on isolated words. The incongruence between exposure in context and testing in isolation may have placed a disadvantage on the semantic condition relative to the non-semantic condition, where exposure and test both occurred in isolation (cf. Martin-Chang et al. 2007). One way to avoid this confounds in future studies would be to present all novel words in contexts that vary in how much meaning they provide for the novel words, or to test orthographic learning not only in isolation, but also in context.

Focussing specifically on the semantic condition, it is possible that semantic learning was too slow to beneficially impact orthographic learning above and beyond the observed levels, despite the fact that the exposure phase was extended over three sessions. Based on the results of Experiment 2, it can be assumed that at the start of session two, semantic knowledge was likely still relatively low. It is unknown how much semantic knowledge was acquired after each session in the present experiments, but it is possible that it took the full three sessions for word meanings to be sufficiently acquired. Once semantic representations are established, they could aid decoding efforts and thus benefit orthographic learning. In the present experiments, performance was generally high on measures of orthographic learning, possibly heading towards ceiling. Thus, it would be interesting to investigate this using a set of novel words with more difficult spellings, or with younger children. Additionally, words with easier meanings than current words could be used. This might result in relatively faster semantic learning and slower orthographic learning, allowing semantic knowledge to emerge in time to influence ongoing orthographic learning.

Another possibility is that both conditions were inherently sufficient to bring about good levels orthographic learning. This could be due to the fact that, at least for the particular samples and sets of words used, exposure to the written forms in and of itself was enough to establish solid orthographic learning, in line with Nation et al. (2007), who found equal learning after reading isolated nonwords and nonwords in story contexts. Reading aloud provides access to or the construction of lexical phonology, which has previously been contrasted with semantics. In a training study with adults using an artificial orthography, Taylor, Plunkett, and Nation (2011) showed a benefit for semantic knowledge on orthographic learning over and above lexical phonology late in the exposure phase for low-frequency inconsistent words. For studies measuring reading accuracy rather than learning, benefits of semantic knowledge over and above lexical phonology were found in 6-year-olds (Ricketts, Davies, Masterson, Stuart, & Duff, submitted), but not 7-year-olds (Nation & Cocksey, 2009). This is consistent with the lack of a semantic benefit in the presence of lexical phonology in the present experiments.

Experiments 3.1 and 3.2 differed in the extent to which lexical phonology was emphasised during the exposure phase. In Experiment 3.1, novel words were read aloud and feedback was given on pronunciations, which provided consistent (and correct) lexical phonology for the novel words. Moreover, in the non-semantic condition, the feedback may have served to highlight that the novel words were indeed unfamiliar English words with lexical status, rather than nonsense words. Experiment 3.2 removed feedback on pronunciations to investigate whether it is indeed critical to the learning and lexicalisation of the novel words in the non-semantic condition. This was not the case, as the results revealed the same pattern as Experiment 3.1, with no difference between semantic and non-semantic conditions in the absence of feedback. The possibility remains that reading aloud the novel words embedded in lists of familiar lists was sufficient to provide lexical phonology and lexical status, even in the absence of feedback. This notion is supported by the fact the reading accuracy at post-exposure test did not differ between conditions, suggesting that children were equally successful in assigning the correct pronunciation to the novel words. Future research could investigate this issue further using silent reading (cf. de Jong & Share, 2007).

In conclusion, in Experiments 3.1 and 3.2, a prime-lexicality effect emerged across conditions, which in adults has only been shown following a training paradigm involving semantics (Qiao et al., 2013). In contrast to the adult findings, both of the present experiments found evidence of children's lexicalisation of novel words regardless of whether semantics were provided. The experiments consistently showed that encountering novel written words in context does not result in superior orthographic learning compared to reading them in isolation without any meaning provided, in line with some existing research (Nation et al., 2007). This was true both in the presence and absence of feedback on children's pronunciations during the exposure phase. The present experiments are the first to show this result for both measures of lexical configuration and lexical engagement.

Chapter 5:

Comparing the Effects of Repetition and Contextual Diversity

Experiment 4

5.1 Introduction

Contextual diversity has become a lexical variable of interest regarding its role in the processing of written words, in particular skilled visual word recognition. It has been contrasted with word frequency, which is based on the number of occurrences of words (e.g. Adelman, Brown, & Quesada, 2006). Unlike frequency, contextual diversity acknowledges that written words are usually encountered surrounded by other words; their contexts or lexical environments. It has been argued that the contextual nature of word occurrences is of importance, whereas frequency counts occurrences irrespective of contexts. Measures of contextual diversity can broadly be divided into simple counts of the *number* of contexts a word appears in, and those that take into account the *diversity* or *variability* of the words that make up those contexts. These measures are usually calculated based on large linguistic corpora and are related to performance based on how much variance they explain in lexical processing in large data sets from adults (for an overview, see Keuleers & Balota, 2015).

The present experiment is based on the argument that research using corpus-derived measures of contextual diversity can inform hypotheses for word learning. Corpus-based measures have the advantage of modelling a much greater number of exposures than what can fit into a laboratory training study. Thus, it can be argued that they potentially come closer to a lifetime experience of words than experiments often spanning just days or weeks. However, unlike training studies that manipulate participants' experiences directly, corpus values can only approximate an individual's experience. Nonetheless, corpus research can point out candidate variables to manipulate in orthographic learning studies based on correlations between corpus variables and performance. Moving beyond correlational evidence, laboratory training studies can in turn provide insights into the mechanisms underlying the relationships observed between lexical variables and skilled performance. The following will discuss evidence for contextual diversity as a valid candidate for word learning research, based on its relationship to skilled word recognition. In this respect, word-learning studies that followed this approach and directly manipulated contextual diversity will be discussed, as well as the design and aims of the present experiment.

5.1.1 Contextual diversity as a predictor of skilled visual word recognition

Contextual diversity has been operationalised in different ways to yield quantifiable measures. It is fair to say that there is no consensus yet as to how diversity should be captured; thus I begin by reviewing the various measures that have been implemented.

Document count. Adelman et al. (2006) measured contextual diversity based on the number of documents a word appears in in a corpus ('document count'). Mathematically, this yields different scores from traditional word frequency counts if words differ in their levels of clustering in the documents they occur in. For instance, two words might appear with equal frequency, but differ in the number of documents across which occurrences are spread. This would yield equal values for word frequency, but different document counts. To demonstrate the importance of contextual diversity for visual word recognition, word frequency and document count values calculated for three different corpora were used as predictors of word naming times and lexical decision latencies from a number of adult studies, covarying for word length, neighbourhood density, rime consistency, number of syllables and initial phoneme. Document count was shown to have unique effects on both naming and lexical decision, with higher document count predicting shorter latencies. There was no unique effect of word frequency once document count was controlled. This was regarded as strong evidence that document count has predictive power beyond what is captured by raw frequency. However, it has to be noted that document count and frequency were highly correlated (r >.95), an issue discussed later on.

Plummer, Perea, and Rayner (2014) contrasted the effects of word frequency and document count on fixation durations of words embedded in sentences, read by adults. Fixation durations were compared for words with similar document count, but higher or lower word frequency, and for words with similar word frequency, but higher or lower document count, based on values obtained from a US subtitle corpus (SUBTLEX, Brysbaert & New, 2009). When word frequency was constant, there was a significant difference between those with higher and lower document counts. Fixation times were shorter for words with higher than lower document counts, indicating a processing advantage for words appearing in a greater number of documents. In contrast, when document count was constant, there was no difference in fixation times between higher and lower frequency words. This suggest again that once document count is taken into account, frequency effects in skilled word recognition may disappear, while effects of document count can be found even when controlling for frequency.

Taken together, these studies suggest that document count has more explanatory power than word frequency. However, by virtue of being count-based, it is conceptually still strongly related to frequency, and the two are highly correlated (r > .95, Adelman et al., 2006). While it might constitute a better measure of the number of a word's occurrences than raw frequency, document count has also been criticised for not taking into account the semantic content/information provided by the documents (Hoffman et al., 2013; Johns et al., 2015). Other measures of contextual diversity have been proposed that capture statistics about the contexts in which words appear.

Contextual distinctiveness. McDonald and Shillcock (2001) proposed contextual distinctiveness as a measure of the linguistic contexts around a given word that captures something beyond their count, in particular, the distribution of co-occurring words. Here, contextual distinctiveness refers to the predictability of contexts around a word. For words that appear in a wide variety of contexts (e.g. 'run'), it is difficult to predict their contexts just based a given word, whereas words that appear in more constrained or specific contexts are more contextually distinct (e.g. 'amok'), i.e. the word itself is more predictive of the kind of contexts it is likely to occur in.

To assess its relationship with lexical processing, McDonald and Shillcock (2001) calculated contextual distinctiveness from the spoken British National Corpus and used the metric to predict lexical decision times from Balota, Cortese, and Pilotti's (1999) dataset, controlling for frequency. Words with higher contextual distinctiveness (e.g. 'run') had longer latencies than words with lower contextual distinctiveness (e.g. 'amok'), i.e. words that occur in more diverse contexts. When controlling for frequency, the correlation between contextual distinctiveness and RT was significant (r = .18). When controlling for contextual distinctiveness, the correlation between frequency and RT was smaller (r = .06), but also significant. This difference in the size of the partial correlations indicated that contextual distinctiveness was a better predictor of lexical decision latencies than frequency. However, as with document count, contextual distinctiveness was overall negatively correlated with frequency (r = -.82); highly frequent words tend to occur in less distinct (i.e. more diverse) contexts.

Semantic distinctiveness. Jones et al. (2012) developed a measure labelled semantic distinctiveness as an improvement over Adelman et al.'s (2006) document count. Like McDonald and Shillcock's (2001) measure, semantic distinctiveness is calculated based on lexical co-occurrence statistics, in this case word overlap between all documents a word appears in. Lexical co-occurrence has previously been shown to be informative about word meanings (e.g. Lund & Burgess, 1996), hence the authors argued that it can be used as a proxy for the semantic overlap between contexts. While McDonald and Shillcock (2001) defined contexts as ten-word windows around words, Jones et al. (2012) calculated semantic distinctiveness was based on entire documents, e.g. newspaper articles or stories. For clarification; Jones et al.'s (2012) semantic distinctiveness variable increases with more variable contexts, whereas McDonald and Shillcock's (2001) contextual distinctiveness variable increases with more similar contexts. Semantic distinctiveness was calculated as the mean dissimilarity between all contexts a word appears in. It was argued that words appearing in semantically distinct contexts suffer from less semantic redundancy, i.e. are encountered in contexts which offer more novel semantic information than similar, overlapping contexts.

To further differentiate words encountered in equally distinct contexts, but differing in the number of contexts, the authors proposed semantic distinctiveness to be derived from a count of all documents a word appears in weighted by their semantic distinctiveness. Semantic distinctiveness values were calculated for words from three different corpora and used to predict lexical decision and naming times from the English Lexicon Project. For words appearing in an equal number of documents, those appearing in contexts with greater semantic distinctiveness were associated with shorter lexical decision and naming times than those appearing in semantically redundant contexts. In a regression analysis, semantic distinctiveness explained unique variance when covarying frequency and documents count, for both lexical decision and naming times. On the contrary, word frequency accounted for little unique variance when covarying semantic distinctiveness and document count. This suggests that the content of a word's contexts matters for word recognition over and above document count and frequency.

Semantic diversity. Hoffman et al., (2013) also developed a measure that aimed to capture the semantic similarity or diversity of contexts a word appears in. Like Jones et al.'s (2012) measure, Hoffman et al.'s (2013) semantic diversity measure was based on word co-occurrences. Rather than using raw word co-occurrence vectors to indicate the extent to which contexts overlap, latent semantic analysis was used to uncover higher-order patterns of co-occurrence. Based on these patterns, semantic distinctiveness of any given word is calculated as the average similarity of the contexts it appears in. As all diversity measures, semantic diversity was positively correlated with frequency (r = .49), based on calculations for 325 words from the British National Corpus. Hoffman and Woollams (2015) related semantic variability to performance on visual word recognition, as measured by lexical decision. Adults were faster on lexical decision for high semantic diversity words than low semantic diversity words. These results were in line with simulations using a connectionist model. High semantic diversity words elicited activation over semantic units more quickly than low semantic diversity words, which could account for the advantage of high semantic diversity words on lexical decision.

In summary, a range of different corpus measures of contextual diversity have been investigated in their relation to visual word recognition. While these differ in their implementation, all of them have been found to capture something different to frequency.

Contextual diversity effects in children's word recognition. Perea et al. (2013) provided preliminary evidence for a role of contextual diversity in children's word reading. They chose Adelman et al.'s (2006) metric of document count, using a factorial design with frequency, to measure lexical decision times in children. Twenty-two 4th graders performed a go/no-go lexical decision task on 60 Portuguese words taken from ESCOLEX, a children's lexical database of elementary textbooks. Document count was calculated based on the number of textbooks a word appeared it. Words were divided into three conditions: high contextual diversity-high frequency, low contextual diversity-low frequency, and low contextual diversity-low frequency. High contextual diversity-low frequency words could not be obtained due to the natural correlation between frequency and diversity. Words were matched on known lexical predictors, including length, imageability, concreteness and number of orthographic neighbours. Children performed a go/no-go lexical decision tasks. Due to the imbalanced design, planned comparisons were performed, revealing faster RT for high contextual diversity-high frequency words than low-contextual diversity-high frequency words. There was no difference between low contextual diversity-high frequency and low contextual diversity-low frequency. A regression analysis revealed contextual diversity and imageability as significant predictors, but not frequency. The results of the regression analysis were replicated in a second set of 27 children with 60 new target words. This can be seen as promising preliminary data that contextual diversity as indexed by document count influences children's processing of written words. However, given the close relationship between document count and frequency, future experiments should use diversity measures that take into account the similarity of contexts, as described above. Furthermore, the use of corpus-based indices of contextual diversity can only provide correlational evidence, and merely estimates the actual reading experience children might have had of these words, and may not apply to all samples. Introducing novel words and manipulating diversity experimentally avoids these issues, which is the approach chosen in the present study.

In summary, the above studies provide good evidence that contextual diversity statistics extracted from corpora explain unique variance over and above word frequency in skilled visual word recognition. Higher ratings of diversity are associated with shorter latencies on visual word recognition tasks. Measures that take into account the semantic similarity of contexts seem to be superior to those relying solely on counting contexts. This is important for theories attempting to explain the underlying conceptual reasons why contextual diversity influences written word processing, as it could point towards a role for the semantic content of a word's contexts.

5.1.2 Does contextual diversity play a role in word learning?

It was not an aim of the present research to provide differential support for one measure of contextual diversity over another. Hence, unless otherwise specified, the term contextual diversity is used as a general concept that is reflected to varying degrees in the specific measures discussed above, but is not tied to a specific mathematical implementation. If one accepts that corpus-derived statistics can be thought of as proxies for an individual's experience, then it appears that the contextual diversity across all exposures to a written word over an individual's life so far, whether child or adult, determines how easily this word is processed now. As argued in earlier chapters, the obvious link between earlier exposure and later processing skill is learning. If contextual diversity influences skilled visual word recognition due to an earlier influence on word learning, this should be supported by evidence demonstrating superior word learning for novel words encountered in diverse contexts compared to non-diverse contexts. Hence contextual diversity is an important candidate for orthographic learning studies. Very few experiments have examined the effects of contextual diversity on learning experimentally using training paradigms. All experiments have been with adult participants; the present experiment makes a start in addressing this lack of developmental research.

Jones et al. (2012) trained adult participants on an artificial language consisting of 12 one-syllable nonwords. Participants viewed three-word ‘sentences’ accompanied by a picture. Frequency (45 vs. 180 exposures) and contextual diversity (one context vs. eight contexts) were manipulated in a 2 x 2 crossed design. Participants were then tested on lexical decision for the new items, which revealed a significant frequency by contextual diversity interaction. For low-diversity words, there were no latency savings for high-frequency words compared to low-frequency words, consistent with the idea that additional exposures in redundant contexts are of limited benefit. For high-diversity words, significant latency savings emerged for high frequency words, consistent with the idea that additional exposures are beneficial, if they appear in novel contexts. An equally valid way to describe the interaction is that a benefit of contextual diversity only emerged at high frequency, while at low numbers of exposures, there was no difference between high- and low-diversity words. This could be taken as an indication that at lower frequencies, contextual diversity is less likely to play a role. While this experiment offers some preliminary support for a role of contextual diversity in word learning, it is difficult to know how artificial language learning compares to learning to read in one’s native language.

Johns et al. (2015) trained adults on 10 nonwords embedded in more realistic, English language contexts. Participants encountered half of the nonwords in five low-variability contexts and the other half in five high-variability contexts. Low-variability contexts concerned a single discourse topic, while high-variability contexts concerned five different topics. These were read silently with the instruction to perform ratings on their clarity and comprehensibility. Following exposure, participants were faster and more accurate on lexical decision for items encountered in high-variability than low-variability contexts. In contrast, on a semantic similarity rating task, participants rated nonwords encountered in low-variability contexts as more similar to close associates than nonwords encountered high-variability contexts, suggesting that semantic learning was better for low-variability contexts than high-variability contexts. These data suggest that contextual diversity might have opposing effects on form and meaning acquisition. This finding is consistent with a similar discrepancy found in skilled processing of words, where an advantage of high contextual diversity was observed for lexical decision, but a disadvantage was found for semantic relatedness judgements (Woollams & Hoffman, 2015). In this respect, Woollams and Hoffman (2015) argued that words occurring in diverse contexts may have noisier, less precise semantic representations, which could explain why performance on semantic similarity ratings was poorer for words encountered in high-variability contexts in Johns et al. (2015). The authors also suggested that the detrimental effect of diversity on semantic learning may have been due to the relatively low number of exposures. Semantic learning from high-variability contexts may require more exposures than low-variability contexts for a core meaning to be abstracted successfully.

Bolger et al. (2008) provided mixed evidence regarding the effect of contextual diversity on orthographic learning. Participants were exposed to rare English words either in the same sentence contexts or four different sentence contexts. Importantly, participants were told that the purpose of the study was to learn new words and that they would be tested on them. Hence the results may not generalise to incidental learning. In Experiment 1, they also manipulated the presence of dictionary definitions in addition to the sentence contexts. In the absence of definitions, both meaning generation and orthographic choice performance were better following varied than repeated contexts. Experiment 2 excluded definitions but introduced a no-context baseline. Participants performed more accurately on meaning generation and were faster at a sentence completion task for words encountered in varied contexts than repeated contexts. However, in this experiment contextual diversity did not affect orthographic choice performance. It is unclear why the effect of contextual diversity on orthographic learning failed to replicate across the two experiments. Furthermore, unlike Johns et al. (2015), this study found a positive effect of contextual diversity on semantic learning, albeit using a task closer to recall, meaning generation, rather than the recognition task (semantic similarity ratings) used by Johns et al. (2015).

No unifying theory has been proposed as to why contextual diversity appears to, at least in some studies, influence not only skilled word recognition, but also the ease with which words are acquired in the first place. Jones et al. (2012) proposed the semantic distinctiveness memory model as a computational simulation of their findings. According to this model, at each encounter with a word, the strength with which it is encoded in memory depends on how much novel semantic information about the word is provided by the context, compared to the existing memorial representation of the word. The results from both Jones et al. (2012) and Johns et al. (2015) were consistent with the simulation results from a computational implementation of the semantic distinctiveness model. However, Johns et al. (2015) clearly state that their model is not intended to be conceptual, suggesting that one should be cautious in using it to derive a theoretical explanation of the effects of contextual diversity on lexical processing.

With regard to semantic learning, Bolger et al. (2008) proposed the 'context variability hypothesis', which proposed that diverse contexts are more favourable to learning about meaning than contexts that do not vary much. This was based on an instance-based memory model of word learning, where each encounter with a word leaves a memory trace. In simplified form, if the same context is encountered repeatedly, the same memory trace becomes reactivated and strengthened, whereas varied contexts leave several memory traces. Importantly, multiple memory traces for the same word allow the abstraction of core features, fragments of which can be reactivated even if the word is encountered in a novel context. Thus, varied contexts are more likely to yield decontextualised knowledge of word meanings, which is more versatile for lexical processing. Unlike the semantic distinctiveness model, this hypothesis is conceptual rather than computational. While Bolger et al. (2008) provided evidence in support of this hypothesis, it does not explain Johns et al.'s (2015) above findings of a negative effect of contextual diversity on semantic similarity ratings. Furthermore, experiments on children's learning of word meanings from storybooks (orally, with spoken language) found either a detrimental effect of varied contexts (Horst et al., 2011) or no benefit of varied contexts over repeated contexts (Wilkinson & Houston-Price, 2013).

Given the overall scarcity of research, it remains unclear whether these discrepant findings are driven by differences in the participant groups, training paradigms, and measures used to assess learning. Only one study has looked at contextual diversity in children's reading (Perea et al., 2013); no study has yet investigated the effects on contextual diversity on children's orthographic learning.

5.1.3 The present experiment

The present experiment examined the effect of contextual diversity on children's orthographic and semantic learning of novel words via incidental exposure in written text. Word learning was compared between a repetition and a diversity condition. Children in the repetition condition read the same stories in each session. Children in the diversity condition encountered words in different set of story contexts in each of three exposure sessions. As such, repetition and diversity conditions differed both in document count (1 vs. 3 different documents per novel word) and diversity between contexts (complete overlap vs. little overlap). Each of the diverse contexts introduced a slightly novel aspect, related to a common core meaning. The number of exposures per word was constant across conditions.

In summary, evidence from adults has shown a beneficial effect of contextual diversity on orthographic processing, as measured by lexical decision. Having established that 12 exposures per novel word are sufficient to drive orthographic learning and induce lexical competition, the present experiment investigated whether variability in the contexts of these exposures results in equal or greater levels of lexical competition. In terms of semantic learning, contradictory evidence made it difficult to propose a definite hypothesis.

5.2. Method

5.2.1 Design

The experiment involved a mixed design with a between-subjects manipulation of exposure condition (repetition vs. diversity). The experiment consisted of three phases – pre-exposure baseline, exposure phase, and post-exposure test. Data for the masked priming lexical decision task were collected before and after exposure, yielding a within-subject factor of test session (pre-exposure vs. post-exposure). Reading accuracy of the novel words was recorded for the first exposure to each word in each of the three exposure sessions. Orthographic decision, spelling and definition knowledge data were collected once, after exposure.

5.2.2 Participants

Twenty children (9 female, 11 male) from Year 5 and Year 6 participated in the experiment. They were aged between 9.17 and 11.75 years (*M* = 10.45, *SD* = .74). Children were selected on the basis of being monolingual native English speakers scoring no more than 1*SD* below the standardised average on the TOWRE (Torgesen et al., 1999). This test consists of two subtests – word reading and nonword reading. Performance on the two subtests was highly correlated (*r* = .79). Children were randomly allocated to either the repetition or the diversity condition, resulting in ten children in the repetition condition and ten in the diversity condition. Descriptive characteristics of the two samples are shown in Table 5.1. The overall mean total standard score on the TOWRE was 110.8 (*SD* = 12.09), indicating that children performed within the population average. Groups did not differ significantly in age or reading ability (see Table 5.1). Children were recruited using an opt-out procedure, where parents were sent two information sheets over the course of two weeks with the opportunity to opt out of their child's participation in the study by returning a response slip.

Table 5.1. Descriptive characteristics of repetition and diversity groups as well as group comparisons (independent-samples t-tests).

	Repetition (N = 10)		Diversity (N = 10)		Group comparison
Measures	*M*	*SD*	*M*	*SD*	*p*
Age (years)	10.68	.83	10.24	.61	.199
TOWRE total raw score[a]	117.5	17.19	113	15.60	.547
TOWRE total standard score[b]	110.9	11.74	110.7	13.06	.972

Notes. [a] total number of words read correctly on both subtests in 45 seconds per test, [b]*M* = 100, *SD* = 15.

5.2.3 Materials and Procedure

5.2.3.1 Word stimuli

The 20 word pairs were identical to those used in previous experiments, each comprising a novel word and an orthographic neighbour base word. For counterbalancing purposes, words were once again divided into lists A and B. Children were allocated to either list based on the order in which they participated in the experiment.

5.2.3.2 Pre-exposure baseline assessment

Masked priming lexical decision. This task was identical to that used in Experiment 2. Due to scheduling constraints, some children completed this task on the same day as the start of the exposure phase, whereas other completed the task on a separate day before the exposure phase.

5.2.3.3. Exposure phase

The exposure phases consisted of three sessions completed on separate days. The number of days across which the exposure phase was completed varied from 3 to 16 days (M = 6, SD = 3.61). This did not differ between the two groups (p = .374).

Story contexts. For the repetition condition, stories were identical to those used in Experiments 2, 3.1, and 3.2. For the diversity condition, these were used for the first exposure session. For the second and third session, new stories were created which differed in narrative plot, but more importantly, also introduced a novel semantic aspect for the target words (see Appendix 5.1). The different aspects all fitted with a common core meaning. For example, the core meaning for ‘undermine’ was ‘to dig under something’. The different semantic aspects introduced by the stories were ‘prisoners digging a tunnel to escape from prison’, ‘moles digging tunnels under a lawn’ and ‘workers digging tunnels to build the London Underground’. The total number of exposures was matched across conditions (Table 5.2) and there was no difference in the mean story length between the repetition condition (M = 163 words) and the diversity condition (M = 162 words). On each day, children read the stories aloud to the experimenter in random order. To keep children interested, in particular those in the repetition condition, filler tasks were created around the sessions. These included caterpillar-themed mazes and Bingo games.

Table 5.2. Number of exposures and contexts for each exposure condition

	Repetition condition	**Diversity condition**
Training Session 1	4 exposures in contexts A	4 exposures in contexts A
Session 2	4 exposures in contexts A	4 exposures in contexts B
Session 3	4 exposures in contexts A	4 exposures in contexts C

Assessment of reading accuracy during exposure. On each of the three exposure days, children's first attempt at reading each of the novel words was scored as either correct or incorrect. Thus, for each day there was one accuracy score for each of the ten novel words per child.

5.2.3.4 Post-exposure assessments

The procedure at post-test was identical to Experiments 2 and 3. Children were tested on definition knowledge, masked priming lexical decision, orthographic decision, and spelling**.** Children were tested between 1 and 5 days after completing the exposure phase (M = 1.89, SD = 1.32), depending on their availability. The test delay did not differ between the two groups (p = .703).

5.3 Results

Analyses were performed using linear mixed effects models following the same principles as earlier chapters. For models with continuous outcome variables, significance was determined using the cut-off point of $|t| > 2$, based on recommendations by Barr et al. (2012). For models with binary outcome variables, mixed effects logistic regression models were computed. Here, the cut-off point for significance was $p < .05$. To control for individual variation in word reading ability, children's total raw TOWRE scores were included as covariates in the analyses.

5.3.1 Reading accuracy of novel words during exposure

Reading accuracy scores were recorded for the first exposure of each novel in each exposure session. Mean scores are shown in Figure 5.1. To compare children's reading accuracy scores on the novel words across between exposure conditions, a mixed effects logistic regression model was computed, as appropriate for binary data. Exposure condition (repetition vs. diversity) was entered as a fixed factor. Models including exposure session as a fixed factor did not converge, so this was not included in the final model (see Appendix 5.2). TOWRE score was a significant covariate (p = .024), showing that children with higher general reading ability read novel words more accurately. The main effect of exposure condition was not significant (p = .790), showing that children in the repetition condition and in the diversity condition read the novel words equally well.

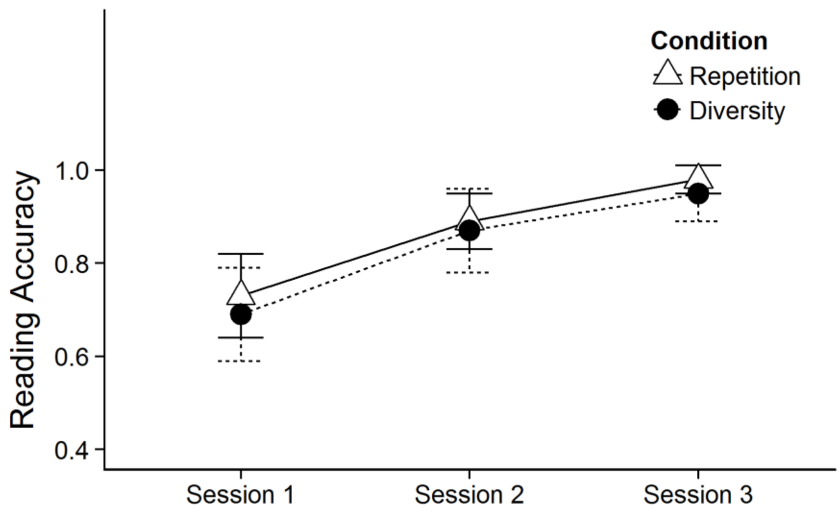


Figure 5.1. Mean reading accuracy scores and 95% CI intervals per exposure session and per condition.

5.3.2 Effects of contextual diversity on novel word learning

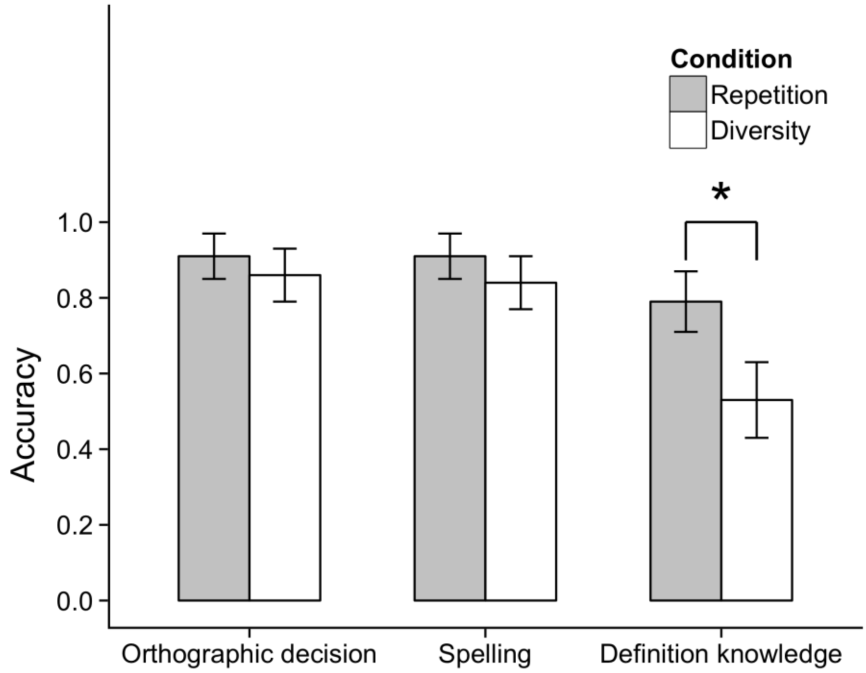


Figure 5.2. Mean accuracy scores per exposure condition for orthographic decision, spelling, and definition knowledge and spelling. Error bars show 95% CIs.

Orthographic learning of novel words was assessed using orthographic decision and spelling. Semantic learning was assessed using definition knowledge. Figure 5.2 shows mean accuracy per exposure condition for each task. To compare performance between exposure conditions, generalised mixed effects logistic regression models were computed for each task (see Appendix 5.2). Exposure condition (repetition vs. diversity) was entered as fixed factor. For orthographic decision, data were missing for one child in the diversity condition. For this task, the difference between exposure conditions was not significant (p = .428), nor was TOWRE a significant covariate (p < .335). For spelling, the maximal model did not converge, so the random effects structure was simplified. The difference between the exposure conditions was not significant (p =

.287). TOWRE was however a significant covariate (p = .038), indicating that children with higher TOWRE scores performed better at spelling than children with lower TOWRE scores.

For definition knowledge, the maximal model did not converge, so the random effects structure was simplified. There was a significant difference between the repetition and diversity conditions (p = .005), with children in the repetition condition recalling significantly more definitions. TOWRE was a significant covariate (p = .018), indicating that children with higher TOWRE scores performed better at definition knowledge than children with lower scores.

In summary, while there was no difference between repetition and diversity condition on orthographic decision and spelling of the novel words; definition knowledge of the novel words was significantly better for the repetition condition than the diversity condition.

5.3.3 Effects of contextual diversity on lexical competition effects on novel word neighbours

Due to an error in the experimental script, data from 40 trials (5%) had to be excluded from the analysis. Overall, children performed with high accuracy on this task, with errors made on only 46 out of 760 (6.05%) trials across sessions. This was deemed insufficient for further analysis. Only trials with correct responses were included in the RT analysis. Based on visual inspection of the qq-plot, RT values were inverse transformed to reduce positive skew, and one data point was removed. Figure 5.3 summaries the data per condition. For ease of interpretation, raw RT data are plotted. Inverse transformed data as used in the analyses are shown in Appendix 5.

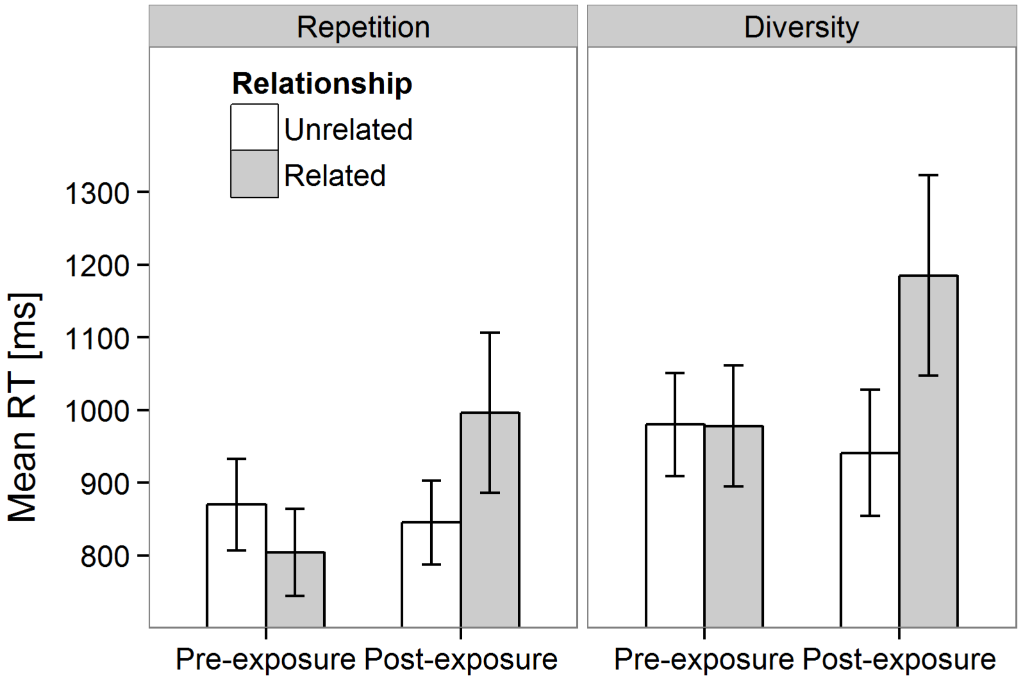


Figure 5.3. Mean RTs for the masked priming lexical decision task, split by exposure condition. Error bars show 95% CIs.

To analyse RT, a linear mixed effect models was computed with exposure condition (repetition vs. diversity), test session (pre- vs. post-exposure), and prime-target relationship (related vs. unrelated) as fixed factors. TOWRE score was not a significant covariate (t = -0.06). The main effects of exposure condition (t = 1.79), test session (t = 0.87), and prime-target relationship (t = 0.53) were not significant. There was however a significant prime-target relationship x test session interaction (t = 3.08, Figure 5.4). There were no further significant interactions (see Appendix 5.2 for full models).

To explore the prime-target relationship x test session interaction, separate models were computed for pre-exposure and post-exposure data, with prime-target relationship as fixed factor. This revealed a trend towards faster responses to targets primed by related novel words rather than unrelated words prior to exposure (t = -1.95). After exposure, children were significantly slower to respond to targets preceded by related primes compared to unrelated primes (t = 2.24). TOWRE score was not a significant covariate either in the pre-exposure (t = 0.78) or the post-exposure (t = -0.39) model.

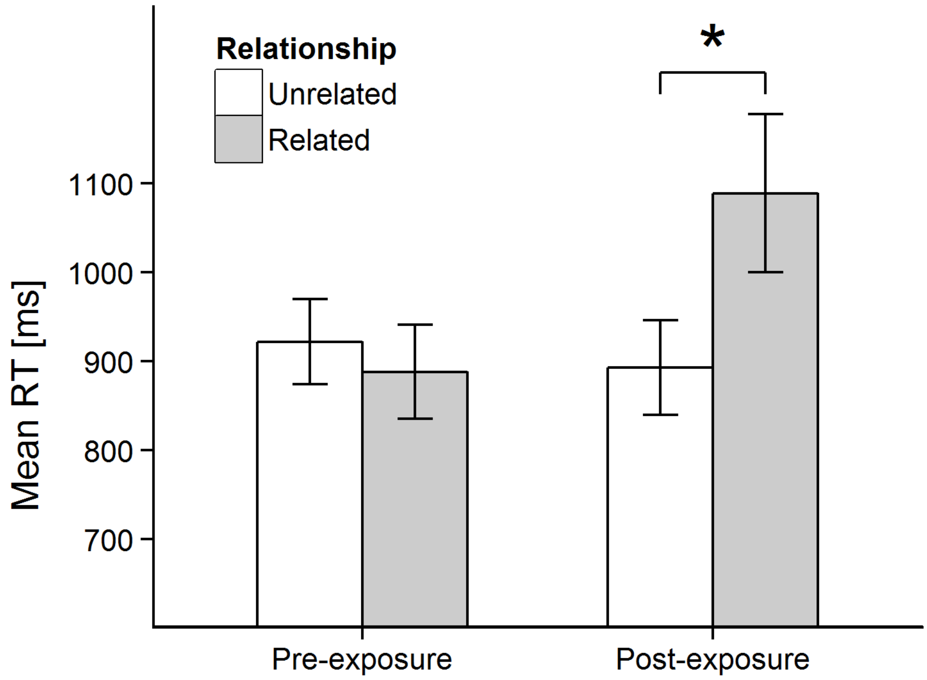


Figure 5.4. Mean RTs for lexical decision task collapsed across exposure conditions, showing the interaction between prime-target relationship and test session. Error bars show 95% CIs.

In summary, the interaction between test session and prime-target relationship indicated a switch in direction of priming effects, from marginal facilitatory priming effects of novel words on their neighbours before exposure to inhibitory priming effects after exposure. Importantly, however, there was no interaction with exposure condition (repetition vs. diversity); both induced the same pattern and strength of priming.

5.4 Discussion

Experiment 4 examined the effects of contextual diversity on orthographic and semantic learning. There was evidence across tasks that contextual diversity did not modulate orthographic learning. Children showed good learning of the novel word forms themselves, as shown by performance in orthographic decision and spelling, irrespective of whether they encountered the novel words in a set of different stories (diversity condition) or repeatedly in the same story (repetition condition). Similarly, lexicalisation of novel words was evident in the emergence of a prime-lexicality effect, which was also not modulated by exposure condition. Prior to exposure, there was a trend towards facilitatory priming by novel words on their neighbours, whereas after exposure, novel word primes significantly slowed responses to their target neighbours. In contrast to learning and lexicalisation of novel word forms, contextual diversity had a negative impact on semantic learning. Children in the diversity condition were less able to define the words than children in the repetition condition.

5.4.1 Effects of contextual diversity on orthographic learning

There was no evidence that encountering novel words in diverse contexts leads to better orthographic learning by children, relative to experiencing them in repeated contexts. This was unexpected given the positive effects of contextual diversity found for word recognition in adults (Adelman et al., 2006; Hoffman et al., 2013; Jones et al., 2012; McDonald & Shillcock, 2001) and children (Perea et al., 2013), and the positive effects found in some studies of orthographic learning in adults (Bolger et al., 2008; Jones et al., 2012). There are a number of potential explanations for this discrepancy.

First, contextual diversity might need to be implemented differently for an effect to emerge. The lack of a well-specified, evidence-based theory explaining the link between contextual diversity and visual word recognition makes it difficult to decide how to implement contextual diversity in orthographic learning studies. Here, contextual diversity was implemented by varying the number of different stories each novel word was experienced in (1 vs. 3). In the diversity condition, the three different stories per novel word involved the same characters and a broadly similar setting (the adventures of a children's gang), but involved different narratives. These narratives introduced a slightly different semantic aspect, as well as novel semantic associations of the novel words, while maintaining a common core meaning. A relationship between contextual diversity and semantic associations was previously demonstrated by Hills, Maouene, Riordan, and Smith (2010), who found a correlation between contextual diversity in caregiver speech and adult word associations. Contextual diversity calculated for words in a caregiver speech corpus was positively correlated with the number of associative cues to which these words were produced (e.g. cat is an associative cue for dog) in adult association norms; words encountered in more diverse speech during childhood were associates of a greater number of words in adulthood. This provides some empirical support for varying the number of semantic associations as a means to implement contextual diversity.

A different implementation would be to systematically vary the extent to which contexts introduce ambiguous meanings or senses. Both a word’s number of associates and its number of senses have been linked to processing advantages in skilled visual word recognition (reviewed in Yap, Pexman, Wellsby, Hargreaves, & Huff, 2012). Hoffman et al. (2013) suggested that the semantic ambiguity of a word's meaning is reflected in the semantic diversity of its contexts. They also demonstrated a positive correlation between their measure of semantic diversity and number of senses. It remains for future research to determine whether the present results would replicate if diverse contexts introduced ambiguous meanings or novel senses rather than associates/features of a common core meaning. This could in turn help to clarify the theoretical basis of the link between contextual diversity and skilled visual word recognition, that is, whether the processing advantage for diverse words is indeed driven by an underlying relationship between a word's semantic characteristics (e.g. number of associates, number of senses, or semantic ambiguity) and the diversity of its contexts.

Alternatively, general learning principles underpin the effects of contextual diversity in orthographic learning and lexical decision, and these are inherently semantic. For example, Jones et al.'s (2012) computational model of semantic distinctiveness is based on the principle that the encoding strength of encounters with a word depends on the novelty of the information provided by their context. Here, the crucial difference between diverse and repeated contexts is in the novelty vs. redundancy of the information provided at each encounter. In contrast, the above semantic accounts argue that contextual diversity effects are (at least partially) driven by semantic differences between words encountered in diverse vs. non-diverse contexts. With regard to the present experiment, the use of a consistent narrative setting across diverse contexts may have reduced the difference in novelty between repetition and diversity conditions, which might explain why there was equal orthographic learning across conditions. Future experiments could vary the degree of novelty at each encounter while keeping semantic factors such as number of associates or senses constant as much as possible. For example, word overlap, narrative settings, or the format in which text is presented (e.g. fiction, non-fiction, or comics), could be varied.

Finally, the lengths of contexts might be important. The present experiment used story-length contexts. Some researchers have implemented contextual diversity in training studies using simple sentences (e.g. Bolger et al., 2008; Jones et al., 2012) or in computational implementations using 10-word windows around a target word (McDonald & Shillcock, 2001). Other computational studies have used longer contexts, more in line with the present study (Johns et al., 2015). Given that there are other significant methodological differences between these studies, further direct manipulation of context length is necessary to investigate its importance.

A quite different explanation of the equivalent learning seen between conditions is that children are robust orthographic learners. Considering the previous experiments reported in this thesis, it appears that orthographic learning via reading aloud stories is robust in children in late primary school. Learning of novel word forms (lexical configuration) was unaffected by all experimental manipulation in this thesis. Similarly, lexicalisation indexed by lexical competition was not affected by explicit attention being drawn to spellings and meanings (Experiment 1), the presence or absence of meaningful contexts (Experiments 3.1 and 3.2) or the diversity of these contexts (Experiment 4). However, in Experiment 2, lexicalisation of novel words differed between the twelve-exposures condition and the four-exposures condition. A prime-lexicality effect emerged after twelve exposures over three sessions, but not after four exposures in one session. It remains to be examined whether this finding was due to the difference in number of exposures or number of sessions. Nonetheless, this evidence is consistent with the idea that as long as children, at least those in late primary school, encounter a novel word form a sufficient number of times (or across sufficient exposure sessions), other factors might have minimal effects on lexicalisation. These results raise the question of when and how effects of variables such as contextual diversity on skilled visual word recognition emerge, given that here, they were not found in children's early orthographic learning processes. This issue will be discussed further in Chapter 6. Overall, the above considerations make it clear that it would be premature to dismiss a role for contextual diversity in children's orthographic learning based on the present null effect.

5.4.2 Effects on semantic learning

Experiment 4 revealed a negative effect of contextual diversity on semantic learning. This is in line with adult findings by Johns et al. (2015), who demonstrated poorer semantic learning for nonwords encountered in diverse vs. repeated contexts. Adults rated nonwords as more similar to close semantic associates after encountering them in repeated contexts compared with diverse contexts. The present results are also consistent with findings from much younger, pre-reading children by Horst et al. (2011), who showed better semantic learning of new phonological forms following shared reading of repeated than varied picture-story books. They are however in contrast to Bolger et al.'s (2008) finding that adults are better at defining words encountered in diverse than repeated contexts, and are faster to respond to diverse words in a sentence completion task.

The previous considerations regarding the implementation of contextual diversity in Experiment 4 also apply to semantic learning. One striking difference between Bolger et al.'s (2008) design and the other studies is that they used one-sentence contexts, while the other studies used longer contexts. These sentences contained less information to be processed, and importantly contained less unrelated filler material. Furthermore, learning was directed, as participants were instructed to learn the novel words. This may have increased the probability of participants forming strong memory traces abstracted from the varied contexts. In contrast to this, in the present study, children in the diversity condition encountered novel words incidentally in rich story contexts, which contained plenty of unrelated novel information. This may have left them with fewer resources for abstracting and memorising the meanings of the novel words. If context length is indeed crucial, the disadvantage of diverse contexts for semantic learning could be hypothesised to disappear or reverse with shorter contexts, or more directed exposure.

It is also possible that more exposures per novel word are required for a positive effect of contextual diversity on semantic learning to emerge. This could be because the abstraction of core meanings is more likely after a greater number of exposures. According to the context variation hypothesis, children in the diversity condition could be argued to have multiple weak memory traces for each word arising from varied contexts, whereas children in the repetition condition could be argued to form a strong single memory trace for each word that is tightly bound to a specific context. These strong memory traces could explain the greater success at the definitions tasks for children in the repetition condition, but this might be based on specific, context-bound memories. In contrast, children in the diversity condition might be more likely to arrive at a decontextualised definition. The context variation hypothesis suggests that abstraction of core meanings from varied contexts occurs via reactivation of 'fragments of decontextualized knowledge' (Bolger et al., 2008, p. 127), a process which might require more exposures than used here. This possibility could be investigated in a future experiment with increased number of exposures per novel word, which could investigate in more detail the nature and quality of the semantic knowledge children gain from repeated vs. diverse contexts, e.g. whether diverse contexts are more likely to yield decontextualised definitions than repeated contexts[8], or are better at applying their semantic knowledge of a word to novel contexts. Further semantic tasks could investigate the richness of children's semantic knowledge, e.g. the number of associates for a novel word. Again, based on corpus analyses the prediction would be that exposure to more diverse contexts provides a greater number of associates for words than exposures to less diverse contexts (cf. Hills et al., 2010).

Finally, it is important to consider potential connections between semantic and orthographic learning, given that there are known semantic influences on word reading (Taylor et al., 2015). The present results are in line with the findings from Experiments 3.1 and 3.2, where concurrent semantic learning of novel words from context did not modulate how well their forms were acquired. Children in the diversity condition performed just as well on measures of orthographic learning and lexicalisation as children in the repetition condition, despite the latter group having learned more about the word meanings. Ideally, the relationship between orthographic and semantic learning should be examined by item-level analyses, investigating the relationship between acquiring the meaning of a specific word and acquiring its form. This will require data from larger samples being exposed to more items.

[8] In a supplementary analysis of Experiment 4, correct definitions provided by children were coded for whether they directly referenced a story context in which they were experienced. 34% of correct definitions provided by children in the repetition condition referenced story contexts compared to 15% of definitions provided by children in the diversity condition. This difference was however not significant (p = .146). This finding should be considered with caution given the low power in the diversity group, where there were fewer correct definitions overall, and because definitions were rated by one person only.

5.4.4 Conclusion

Experiment 4 did not find an effect of a contextual diversity manipulation on orthographic learning, i.e. learning was good across both conditions. This is at odds with the well-documented positive relationship between contextual diversity and skilled visual word recognition, as indexed by lexical decision. Several possible reasons for this apparent discrepancy were discussed. Given the lack of other training studies investigating this in children, the results have to be considered preliminary and in need of further research. Nevertheless, it is clear that diversity is not required to drive lexical competition, as measured by the prime-lexicality effect. In terms of semantic learning, there was a negative effect of contextual diversity, which is in agreement with some findings with adults (Johns et al., 2015) and much younger children learning spoken words (Horst et al., 2011). However, positive effects have also been found in adults (Bolger et al., 2008). The relatively low levels of semantic learning in the diversity condition in the present experiment could be explained by the large amount of irrelevant contextual information to be processed alongside the novel words, or the difficulty of abstracting a core meaning from varied contexts based on a relatively low number of exposures per novel word. These suggestions could be tested in future experiments.

Chapter 6:

General Discussion

6.1 Overview of Main Findings

This thesis aimed to elucidate children's learning of novel written words via reading experience. In particular, incidental learning via story reading was examined, addressing both methodological and theoretical issues stemming from the existing literature. This final chapter summarises the major findings, as well as setting them in the broader theoretical context. Methodological issues and limitations are discussed, and throughout, issues for future research are identified.

Broadly, the present research assessed children's learning of novel written words in terms of both lexical configuration and lexical engagement, a distinction originating in spoken language research (Leach & Samuel, 2007). Existing research in children's orthographic learning has relied on measures of lexical configuration only, following the self-teaching tradition (e.g. Share, 2004), whereas research in children's acquisition of spoken words has also investigated lexical engagement (e.g. Henderson et al., 2012). The present experiments were the first to directly contrast measures of lexical configuration (orthographic decision and spelling) and lexical engagement (lexical competition paradigms) in children's orthographic learning. Across five experiments, a novel adaptation of the prime-lexicality effect on masked priming lexical decision proved to be an effective measure of lexical engagement. Second, the experiments in this thesis examined a number of theoretical factors of interests, assessing their influence on the above measures. These factors were identified based on findings from research on the self-teaching hypothesis, spoken language acquisition, and skilled visual word recognition.

6.1.1 The prime-lexicality effect as a novel measure of orthographic learning

The present research established the prime-lexicality effect observed in masked priming lexical decision as a promising measure of lexical engagement in children, evidenced by the fact that the effect emerged in all five experiments. In Experiment 1, it emerged across reading-only and reading plus conditions. Using a different set of story materials, the effect also emerged in Experiment 2's twelve-exposures condition. In Experiments 3.1 and 3.2, it emerged across semantic and non-semantic conditions, and in Experiment 4, it emerged across repetition and diversity conditions. By combining the prime-lexicality effect with traditional methods, this research addressed the need for novel, sensitive measures of orthographic learning. First, measures such as orthographic choice and spelling do not by themselves differentiate between lexical configuration and lexical engagement. This distinction was first made in the spoken language literature (Leach & Samuel, 2007), but also provides a useful framework for research in orthographic learning. Using both types of measure, the present experiments approximated how novel written words move from initial episodic familiarity to integration in the lexicon. Assessing orthographic learning comprehensively allowed the detection of differential effects of variables of interest on lexical configuration vs. lexical engagement, as found for the effects of number of exposures in Experiment 2.

Second, the prime-lexicality effect was investigated as a measure of gradation in learning and lexicalisation. In Experiment 2, the effect emerged in the twelve-exposures condition, but not the four-exposures condition. It remains to be seen whether it is possible to move beyond this 'present vs. absent' distinction to more fine-grained differences in the degree of novel-word lexicalisation. In order to detect relative differences in the magnitude of the effect, larger sample sizes than those used in the present experiments will be necessary. Individual differences in the prime-lexicality effect have been successfully demonstrated in a sample of 100 university students (Andrews & Lo, 2012). However, the comparison was made between nonword and high-frequency word primes, rather than between novel words before and after exposure. Hence, while it appears to be possible to detect gradual differences in the magnitude of priming effects, this has yet to be demonstrated in a novel-word training paradigm. In summary, the present research successfully employed the prime-lexicality effect as novel measure of lexical engagement to examine a number of experimental manipulations, discussed in more detail below.

6.1.2 Which factors influence learning of novel written words?

Incidental vs. directed exposure. Experiment 1 compared a reading-only and a reading-plus condition, with the aim of investigating the impact of incidental vs. directed exposure on word learning. Children in both conditions read aloud stories containing novel words. In addition, children in the reading-plus condition wrote out definitions for the novel words and practiced spelling them. Somewhat unexpectedly, there was no difference between the two conditions in the lexical configuration of either form or meaning of the novel words. This could be explained by the overall high performance across tasks, which could have masked difference between conditions. Similarly, a prime-lexicality effect emerged that was not modulated by exposure condition, indicating that words were lexicalised both after incidental and directed exposure and induced competition effects on their neighbours. This finding constituted an important first demonstration of the prime-lexicality effect in a training study with children. In contrast, lexical competition effects were not observed on a novel semantic relatedness judgements task. This task has not previously been used to demonstrate lexical competition, and may thus have been insufficiently sensitive. Hence the following experiments focussed on the prime-lexicality effect to measure lexical engagement.

The experiment also established story reading as powerful opportunity for both orthographic and semantic learning of novel words. However, there were concerns regarding the extent to which the story materials met more stringent criteria for incidental learning, e.g. as set out by Swanborn and DeGlopper (1999). This may have inflated learning in the reading-only condition. This concern was addressed in subsequent experiments, which used a new set of stories designed to be closer to children's natural reading experience. In Experiment 2, children showed good orthographic and semantic learning when reading a new set of stories repeatedly across three sessions, at levels comparable to Experiment 1. This supports the conclusion that incidental exposure to novel words in stories does indeed yield good learning and lexicalisation.

Number of exposures. Experiment 2 investigated effects of number of exposures on children's learning of novel written words and their lexicalisation. In self-teaching experiments, children show orthographic learning after only a few exposures, but this has only been investigated for measures of lexical configuration (e.g. Share, 2004). In contrast, adult studies on lexical engagement of novel written words have shown a comparatively slower time course (e.g. Qiao & Forster, 2013). Experiment 2 investigated the time course of both lexical configuration and lexical engagement of novel written words in children. To do so, two exposure conditions were compared: four exposures per word in one session vs. twelve exposures per word across three sessions. As predicted, lexical configuration of novel word forms, indexed by orthographic decision and spelling, was evident after just four exposures. In contrast, lexical engagement of novel word forms, indexed by the prime-lexicality effect, only emerged after twelve exposures. The results from the twelve-exposures condition replicated the prime-lexicality effect found in Experiment 1. Learning of meaning was also better after twelve exposures than after four exposures, providing an indication that lexical configuration of meaning via incidental exposure may be slower than configuration of form. Overall, the findings were consistent with both Leach and Samuel's (2007) proposal of two different cognitive processes involved in word learning, and the Complementary Learning Systems theory (McClelland et al., 1995). Further research is necessary to understand the basis of these findings: the twelve-exposures condition differed both in terms of the number of exposure, and the number and spacing of exposure sessions from the four-exposures condition. Nonetheless, these findings illustrate the obvious, but important link between reading experience and developing an orthographic lexicon. Children who do not read much are likely to have both fewer exposures to novel words overall, and fewer reading sessions in which they encounter the same words repeatedly. As evident in Experiment 2, this likely leaves them with fewer chances of successful lexicalisation of novel words. In turn, this could have potentially wide-ranging consequences extending to reading comprehension, as 'reading comprehension rests to a considerable extent on knowledge of words' (Perfetti & Hart, 2002, p. 189).

Learning words with meaning vs. no meaning. Experiments 3.1 and 3.2 both compared a semantic (reading stories) and non-semantic condition (reading word lists) in order to compare new word learning in the presence or absence of semantic information. As expected, in both experiments, children in the semantic condition showed good semantic learning of the novel words; they successfully instantiated meaning for the novel words based on the story contexts. In the non-semantic conditions, semantic learning was low; again this was expected as the words were selected to be novel to the children, and little meaning was provided via reading the words in lists.

In terms of orthographic learning, Experiment 3.1 revealed no difference between the non-semantic and semantic condition for both measures of lexical configuration and lexical engagement. Across conditions, children showed good knowledge of the novel word spellings, and their lexicalisation was evident in the emergence of a prime-lexicality effect. The finding of a prime-lexicality effect in the non-semantic condition provides some indication that exposure to form alone might suffice to drive lexicalisation. However, while the word lists did not provide meaning for the novel words, children were nonetheless likely to consider them unfamiliar words with lexical status rather than nonsense words, given that they were read aloud, embedded in lists of familiar words, and lexical phonology was provided consistently by feedback on pronunciations. Experiment 3.2 investigated whether the findings from Experiment 3.1 would replicate in the absence of feedback, which might reduce the novel words' lexical status and emphasis on lexical phonology. The results revealed a pattern of results identical to Experiment 3.1. Children in the non-semantic condition showed novel word learning and lexicalisation to an equal extent as those in the semantic condition. Given that semantics did not seem to be crucial, lexical phonology and lexical status may have driven learning and lexicalisation, even in the absence of feedback on pronunciations, as novel words were read aloud embedded in lists of familiar words. Future research could aim to reduce these factors further to pin down their importance for orthographic learning and lexicalisation, e.g. by using silent reading (cf. de Jong & Share, 2007).

The converging results of Experiments 3.1 and 3.2 raise the theoretical question of how and when semantic effects on visual word recognition arise (e.g. semantic richness effects in lexical decision and word naming, reviewed in Yap et al., 2012), given that they do not seem to affect initial orthographic learning of novel words. This issue will be discussed in more detail below.

Repeated vs. diverse contexts. As the final investigation of children's word learning via reading stories, Experiment 4 examined the nature of the story contexts themselves by varying whether children experienced the same stories repeatedly, or encountered each novel word in three different story contexts. This manipulation was based on research into the role of contextual diversity in skilled adult word recognition (e.g. McDonald & Shillcock, 2001), children's word recognition (Perea et al., 2013), and the learning of novel written words by adults (e.g. Bolger et al., 2008, Johns et al., 2015). The experiment revealed contrasting findings for orthographic vs. semantic learning. Orthographic learning was not affected by the manipulation of contextual diversity. Across conditions, good learning was evident on orthographic decision and spelling, and lexicalisation was equivalent, indicated by the emergence of a prime-lexicality effect in both conditions. Once again, this raises the issue of how and when contextual diversity effects on skilled visual word recognition emerge, if not via their influence on children's learning early in their experience of particular words. In contrast to this, semantic learning of the novel words was significantly better after exposure to repeated story contexts than diverse story contexts. The differences found between orthographic and semantic learning are discussed in more detail below, alongside suggestions for future research.

6.2 Theoretical Implications

6.2.1 Robust orthographic learning vs. variable semantic learning

Based on the above experiments, it can be concluded that orthographic learning is generally robust in 9-11 year-old children with average reading ability. Lexical configuration of novel word forms, assessed using measures in the self-teaching tradition, was good across experiments and not affected by any of the experimental manipulations. In contrast, as predicted, lexical engagement, assessed by the prime-lexicality effect, required more exposures (or sessions) than lexical configuration (Experiment 2). However, if these were given, the extent to which exposure was incidental vs. directed (Experiment 1), the presence vs. absence of context (Experiment 3.1 and 3.2), or the diversity vs. repetition of contexts (Experiment 4), had little impact on lexicalisation outcomes. As such, it appears that factors known to influence skilled visual word recognition in adults (e.g. contextual diversity) do not necessarily affect the early process of acquiring this word recognition ability in the first place. This issue will be discussed in more detail later on.

It is possible to question whether the above findings of robust orthographic learning would generalise to different items and/or participants. Performance on measures of word learning was generally high across tasks and experiments, potentially approaching ceiling, which may have masked subtle effects of the experimental manipulations. If this was the case, such effects may become detectable if task difficulty was increased, i.e. by using a larger set of items to be learned, more difficult words, or younger children. Apart from the issue of measurement sensitivity, there are also theoretical reasons why orthographic learning might be less robust for different items or children.

With regard to items, the present experiments focussed exclusively on words which children were not expected to know in oral vocabulary, as children are known to acquire new words via reading (Nagy et al., 1985). However, in the literature on visual word recognition and its development, it is often assumed that the phonology and semantics of words are already established prior to their first written encounter. For example, in Harm and Seidenberg’s (2004) simulation of word recognition in the triangle model, the model was pretrained heavily on phonology and semantics. The same applies to Ziegler et al.’s (2014) implementation of self-teaching in a CDP model, which was equipped with an existing phonological lexicon. Manipulations of reading experience affecting the provision meaning or phonology, e.g. via context or feedback on pronunciation, may affect orthographic learning differently for words familiar in oral vocabulary than for entirely novel words. For entirely novel words, exposure via reading experience may provide previously unknown semantic, phonological, and orthographic information. In contrast, for words in oral vocabulary, semantic and phonological information may already be familiar. However, semantics and phonology provided via reading experience may not always be compatible with a child’s knowledge in oral vocabulary. For example, for irregular words, phonological decoding attempts may result in an incorrect phonological form, a problem that could be solved via meaningful context (cf. Wang et al., 2011) or feedback provision. In this case, it has already been demonstrated that context affects learning of irregular, but not regular words (Wang et al., 2011). Semantic information provided by context could also be incompatible with children’s oral language knowledge of a word, i.e. in the case of ambiguous words or unfamiliar word senses, which may affect orthographic learning. It remains for future research to directly compare orthographic learning of words unfamiliar vs. familiar in oral vocabulary.

With regard to participants, children in the present experiments were proficient, experienced decoders. General word reading ability was assessed using the TOWRE. As a covariate, TOWRE was not always significant, but as expected, the direction of effects consistently indicated that better readers scored higher on word learning measures and showed shorter latencies on lexical decision. Unfortunately, interactions of word reading ability with experimental manipulations could not be investigated due to lack of variation and small sample sizes. Thus it remains to be investigated whether the results of the manipulations reported here generalise to younger or less proficient readers. Throughout reading development, the orthographic system as a whole is known to evolve, and several markers of developmental change have been suggested. As outlined in Chapter 1, Ehri's (2005, 2014) phase account of the development of visual word recognition proposes a change in the predominant type of connections formed between unfamiliar word forms and phonology, i.e. from individual letters to larger chunks of letters. Compatible with this idea, it has been suggested that children shift from serial decoding of letters to sound to parallel processing of letters for both words and nonwords. Van den Boer and de Jong (2015) suggested that if children's word and nonword reading is predicted by discrete digit naming rather than serial digit naming, they are likely engaged in parallel processing of letters rather than serial decoding. Lastly, Chapter 1 also discussed lexical tuning as a marker that differentiates word recognition processes in beginning vs. skilled readers (Castles et al., 2007). These markers could be used to investigate predictors of orthographic learning, as well as to establish if learning is less robust earlier in reading development, i.e. affected by parameters of children's reading experience manipulated in my training studies.

Beyond reading ability as a predictor of orthographic learning, the relationship likely also exists in the opposite direction. High-quality orthographic knowledge of specific words, sometimes referred to as print vocabulary, is argued to be essential for fluent text reading, and to allow focus on comprehension (e.g. Ehri, 2014; Perfetti & Hart, 2002). The present research focused on the development of print vocabulary via the acquisition of word-specific knowledge, i.e. on how exposure to a given word leads to knowledge and lexicalisation of this word only. In particular, learning of words with previously unknown meanings was investigated. While this revealed evidence of robust orthographic learning, the question remains of how this item-specific process contributes to the development of children's word recognition overall, e.g. generalises to the recognition of other known words as well as unfamiliar words. As described earlier, over the course of reading development, children have been shown to undergo lexical tuning in masked form priming (Castles et al., 2007) and move from serial to parallel letter processing of words as well as nonwords (van den Boer & de Jong, 2015, see also Grainger & Ziegler, 2011). The role of item-level word learning via reading experience in driving these developmental shifts could be investigated longitudinally, using markers such as length effects and lexical tuning as outcome variables, in addition to traditional assessments of reading ability. Regarding the underlying mechanisms, computational models of word recognition could further examine how learning of specific words via print exposure may yield general developmental shifts in how children read words (see also Ehri, 2005, 2014).

Turning now to semantic learning via incidental exposure, this was more variable than orthographic learning. In particular, it was affected by number of exposures/sessions (Experiment 2) and contextual diversity (Experiment 4). A few caveats have to be mentioned regarding this apparent discrepancy between orthographic and semantic learning. The primary focus of the present experiments was on the learning and lexicalisation of novel word forms rather than their meanings. Hence, semantic learning was assessed using a single measure (definition knowledge), which was scored on a binary scale, providing a stringent, but potentially crude measure of semantic knowledge. Thus, an in-depth discussion of the nature of semantic learning and the factors influencing it is beyond the scope of this thesis. However, some considerations can be made regarding the contrast between orthographic and semantic learning. Again, one has to be cautious in making direct comparisons of absolute levels of learning, given that these were measured in different ways. Nonetheless, in line with the present results, semantic learning via incidental exposure is generally harder than orthographic learning. Unlike orthographic learning, semantic learning from context requires inferences and abstraction, and the probability of learning a word from a single exposure in context as measured by definition knowledge has been estimated as just .11 (Nagy et al., 1985). Furthermore, unlike the orthographic form of an English word, which is made up of an orthotactically constrained combination of 26 potential letters, the scope of word meanings is much less constrained.

The differences in orthographic and semantic learning observed here have potential implications for the development of lexical quality, at least for unfamiliar words encountered incidentally via reading experience. The case of words which are already known in oral vocabulary when they are first in encountered in writing might be quite different. For entirely novel words encountered via reading, it appears that the semantic constituent of lexical quality may develop more slowly than the orthographic constituent, and may be affected to a greater degree by number of exposures (Experiment 2) and contextual diversity (Experiment 4). It remains to be investigated to what extent the development of semantic and orthographic constituents interact. For example, in a scenario where children have acquired orthographic knowledge following an initial encounter with a novel written word, a high-quality orthographic representation may allow processing savings during subsequent encounters, compared to a lower-quality orthographic representation. In turn, this might allow semantic knowledge to be acquired more successfully. Such an investigation would require careful tracking of online reading processing and offline learning across multiple exposures, e.g. by measuring fixation durations of eye-movements throughout exposure sessions.

6.2.1 Effects on learning vs. activation

While orthographic learning was robust in the present experiments, many factors are known to influence skilled visual word recognition of familiar words (e.g. Balota et al., 2004). For example, contextual diversity has been shown to predict lexical decision times (e.g. Perea et al., 2013), but was not found to affect orthographic learning in Experiment 4. Whether this inconsistency is surprising is a question in its own right. For instance, one could argue that orthographic learning reflects the establishment of new orthographic representations, whereas skilled visual word recognition reflects the activation of existing representations, and that these two processes should not to be confused. However, it is possible to argue against this, largely because it is difficult to separate learning and experience. This is particularly true for lexical variables that are based on written experience of words, as this experience is also the foundation of orthographic learning. For example, written word frequency counts predict skilled lexical decision times in skilled readers (Brysbaert & New, 2009) and fixation duration in children (Joseph, Nation, & Liversedge, 2013). Counts based on written corpora are intended to serve as a proxy for an individual’s experience of words, which in turn arguably reflects their lifelong learning history. Different types of learning history are precisely what the present laboratory training studies of orthographic learning modelled. From this point of view, frequency effects should also be found when the focus is on learning, provided the laboratory experience comes close enough to real life experience. This was indeed the case for the frequency manipulation in Experiment 2. However, the same did not apply to semantics (Experiments 3.1 and 3.2) or contextual diversity (Experiment 4), where effects found in skilled visual word recognition were not found to affect orthographic learning. It is possible that some variables affect visual word recognition after a greater number of cumulative exposures via reading experience than modeled in the present experiments. Furthermore, oral language experience could also provide an additional influence on visual word recognition. In the present experiments, only words unfamiliar in oral language were considered, and children did not receive any prior spoken language exposure to the novel words. Conceivably, many words are experienced in both written and spoken language, and variables such as semantic knowledge of a word are in many cases likely determined by both types of experience. To summarise, in order to experimentally investigate the emergence of effects on skilled visual word recognition, item-level learning should be tracked over a greater number of exposures, i.e. beyond the point of initial orthographic learning as indexed by spelling or orthographic decision. The influence of previous or subsequent oral language experience could also be contrasted with purely written experience. For an experiment spanning a long period of time, it may be necessary to use nonwords to ensure words are not encountered outside the laboratory, i.e. receive uncontrolled experience. However, this would compromise on close resemblance to natural experience, highlighting the trade-offs between using words and nonwords, discussed in more detail later on. Connectionist models of visual word recognition could also be used to investigate these factors, and their time course during learning.

While the above illustrates the difficulty in taking the ‘learning’ out of the ‘activation’, it is equally problematic to take the ‘activation’ out of the ‘learning’. The activation of existing orthographic knowledge of a novel word form can only be ruled out during its very first encounter. Learning, however, is likely to continue over future exposures to the word, at which point the activation of existing (to be improved) orthographic representations and their *learning* became more difficult to disentangle. This does not only apply to a word's orthography, but conceivably also to phonology and semantics. A first encounter with a novel word could supply some phonological and semantic knowledge, which could affect subsequent encounters. It then becomes a challenge to tease apart their effects on word recognition attempts, and on the orthographic learning process.

6.3 Methodological Considerations

6.3.1 Using novel words vs. nonwords

Children frequently encounter unfamiliar words in their everyday reading experience (Nagy et al., 1985). The present experiments aimed to closely imitate this scenario by using rare English words, which based on pilot data were unfamiliar to 9-11-year-olds. Alternatively, nonwords could have been used to rule out any possibility of children knowing the words prior the experiments. However, nonwords often appear artificial and thus may be perceived as nonsense or made-up words. It is also possible to raise ethical concerns regarding the practice of teaching nonwords to children. The trade-off when using word stimuli in order to resemble children's real-life experiences is reduced experimental control over lexical variables. In the present experiments, items were particularly constrained by the requirement of finding pairs of high-frequency basewords to act as targets (e.g. *field*) and low-frequency neighbours to be trained and act as primes (e.g. *wield*), with the additional constraint that the words should not have additional high-frequency neighbours. The resulting stimulus set was relatively heterogeneous, e.g. regarding word class. Stimulus matching is particularly important when sets of items are compared which are selected to vary for a particular variable of interest (e.g. contextual diversity), but are otherwise matched for other lexical variables which could influence processing (e.g. frequency, imageability). Stimulus matching problems can be overcome by comparing a word with itself (see Bowers et al., 2005, for further discussion). For the prime-lexicality effect, this was achieved in the present experiments by comparing pre-exposure and post-exposure performance on the same items for each child, i.e. the same counterbalanced list. Regarding the comparison between exposure conditions (e.g. semantic vs. non-semantic), these were made between-subjects, i.e. using the same two counterbalanced lists in both conditions, rather than comparing different words within-subject.

Although words, the novel words had nonword status prior to exposure. For the present experiments, words were piloted to ensure that they were unlikely to be known by children aged 9 to 11. This was also confirmed by Experiments 3.1 and 3.2, where children in the non-semantic condition showed very little definition knowledge of the novel words following exposure. This is important given that a number of lexical variables are driven by item-specific experience, so by limiting the likelihood of prior experience of the novel words, these can be controlled. Assuming children encountered the novel words for the first time during the experiments, variables such as frequency and age-of-acquisition were determined by the experiments. Frequency can be assumed to be equal to the number of times children experienced words during the experiment, and age of acquisition, in the case of the acquired word, was their age at participation. Semantic variables such as imageability and concreteness were likely determined by the story contexts, which may have led to variability between items, but again, words were compared with themselves. The same applies to orthographic and phonological variables such as phonological complexity, sound-spelling transparency and letter position.

In principle, there should not be a difference between nonwords and unknown words the in prime-lexicality effect, because in both cases facilitatory priming is believed to be driven be orthographic overlap in the absence of lexical competition. By definition, neither nonwords nor unknown words are in the lexicon, so both should show facilitatory priming for their orthographic neighbours. However, differences are possible with regard to learning and lexicalisation. As argued above, learning nonwords in the laboratory may not be representative of how unfamiliar words become familiar via every-day experience.

6.1.3 Conducting training studies with children

Theory-based training studies are invaluable for testing causal hypothesis regarding the influence of factors of interest on reading development (Hulme & Snowling, 2014; Snowling & Hulme, 2011). They do however pose methodological challenges. The present research involved time-consuming training regimes that are difficult to implement logistically in schools, meaning that a few methodological compromises were made. First, some of the data collected for this thesis were used for comparison of exposure conditions in multiple experiments. For Experiments 2, 3.1, and 4, an experimental comparison was made against the same baseline condition involving story reading across three sessions (the twelve-exposures condition in Experiment 2, the semantic condition in Experiment 3.1, and the repetition condition in Experiment 4). Rather than re-collecting data for this baseline condition, the same data was used as the baseline condition in these experiments. This approach was taken in order to maximise the number of theoretical questions that could be addressed in the available time.

However this introduced potential issues due to the fact that data for different conditions were collected at different time points and in different schools. In order to control for potential differences between the samples of children in each condition, reading ability as measured by the TOWRE was controlled in all analyses, and TOWRE scores were generally similar across experiments and conditions. However, if the data in the baseline condition were for some reason not reliable, this would carry through all of the experiments. This possibility is unlikely given that the prime-lexicality effect was nonetheless replicated for an entirely different sample and different materials in Experiment 1, and for a different sample but identical materials in Experiment 3.2. Furthermore, an equivalent prime-lexicality effect also emerged across conditions in Experiment 3.1, which included newly collected data for the non-semantic condition, and in Experiment 4, which included newly collected data for the diversity condition. Hence, it can be argued that the prime-lexicality effect was replicated sufficiently across experiments, despite the repetition of baseline data in some cases.

A second issue is sample size. For comparisons across exposure conditions, sample sizes per condition varied between 10 and 15 children, which is arguably small. The number of novel words experienced by each child was also limited to 10, given that they were embedded in relatively lengthy story contexts. This might explain why there was variability between experiments in the extent of facilitatory priming prior to exposure, which was non-significant in Experiment 1, marginally significant in Experiment 2, 3.1, and 4, and significant in Experiment 3.2. In the adult literature, this difference is numerically small, but significant (e.g. ranging from 14-30 ms in Qiao et al., 2009). Given that children's behaviour is likely more variable and affected by noise, larger samples might be required than in adult experiments. Given that all five experiments in this thesis used masked priming lexical decision as the first task administered in the pre-exposure session, these data could be combined across experiments. Data from 89 children were included in the analysis. Controlling for TOWRE score, children were significantly faster at lexical decisions to targets primed by related novel word primes (e.g. *wield - FIELD*) than unrelated control primes (t = -3.22). This facilitatory priming effect indicates that novel words behaved like nonwords prior to exposure, as in adult experiments. TOWRE was a marginal covariate, indicating a trend towards children with higher scores showing overall faster responses than children with lower scores (t = -1.90, see Appendix 6 for full model). Demonstrating this facilitatory priming effect prior to exposure in a larger sample is important, because it provides evidence that children show priming effects as expected for nonwords based on adult experiments (e.g. Davis & Lupker, 2006). This is an important prerequisite for using the prime-lexicality effect as a measure of learning in children. There were also differences between experiments regarding whether lexical competition after exposure was indicated by a lack of facilitatory priming, or the emergence of inhibitory priming. However, an exploratory analysis of the post-exposure data from 80 children combined across experiments[9] revealed significant inhibitory priming (t = -3.13, see Appendix 6 for full model). This should be considered with caution given the differences between exposure conditions[10].

As a general caveat, it has to be noted that the use of portable laptops in school settings was not ideal for the very precise detection of priming effects, due to small amounts of variability in the duration of stimulus presentation. In instances where the prime presentation was delayed slightly beyond the ideal 57 ms, priming effects may

[9] Data from the four-exposures condition in Experiment 2 were not included, as a prime-lexicality effect did not emerge here.
[10] Exposure condition was included in the model as a random effect. TOWRE was a significant covariate,(t = -4.52), indicating longer RTs for children with lower TOWRE scores than those with higher scores.

have been inflated. However, any general issues with timing and presentations would be equivalent across exposure conditions, and for pre-exposure vs. post-exposure. For the latter, the prime-lexicality effect, indicated by a switch in the direction of priming, was replicated across all five experiments.

A further potential concern is the variation in length of the exposure phase and test delay between children. Rather than adhering to a strict schedule (e.g. an exposure phase across three consecutive days with a test after 24 hours), a flexible approach was taken. A strict schedule would have invariably led to exclusion of a large proportion of data, as children might miss sessions due to absence or constraints set by their school. Thus, there was variation between children in the spacing of exposure sessions, which could have affected learning. Effects of spaced vs. distributed practice on novel word learning have been investigated, in particular for the acquisition of meaning. For example, Sobel et al. (2007) fount that fifth graders showed superior long-term retention of novel word meanings following two spaced learning sessions (1-week interval) than two massed learning sessions. The role of spacing in orthographic learning remains an important subject for future research. Similarly, systematic manipulation of the delay between exposure and test could be used to investigate the durability of learning and lexicalisation. Previous research has shown better orthographic learning in children learning to read in English when tested one day after exposure, relative to a 7-day delay (Nation et al., 2007). However, learning was only assessed using measures of lexical configuration, e.g. orthographic choice, which may be supported mainly by temporary episodic memories. The durability of children’s lexicalisation of novel word forms, as measured by the prime-lexicality effect, remains to be investigated. Furthermore, it is possible that some of the factors investigated here (e.g. semantics or contextual diversity) would show an influence after longer delays, allowing for time for consolidation.

6.3.3 Episodic interference vs. lexical competition accounts of the prime-lexicality effect

In this thesis, the prime-lexicality effect was used in a novel way, as a measure of orthographic learning in children, with the purpose of elucidating the factors that influence learning. As such, the aim was not to explore the prime-lexicality effect itself. Nevertheless, it is important to consider what it reflects. As discussed in Chapter 1, the prime-lexicality effect is considered a marker of lexicalisation (e.g. Qiao et al., 2009), because it distinguishes between nonword and words via their characteristic priming effects. Inhibitory priming between words and their neighbours is thought to be driven by lexical competition between entries in the lexicon; by definition, nonwords are not in the lexicon and are thus not available to inhibit related words.

Qiao et al. (2009) introduced the prime-lexicality effect as a measure of lexicalisation in order to address their concern that interference effects found on semantic categorisation (artefact vs. natural object, Bowers et al. 2005) were driven by episodic interference/biases rather than lexical competition. They argued that the repeated typing task used by Bowers et al. (2005) may have created a strong episodic trace for nonwords such as *banara*, which may have given rise to a post-access spelling check in order to discriminate *banana* and *banara*, rather than lexical competition (Qiao et al., 2009). They argued that episodic interference alone is not evidence for lexicalisation of novel words. In contrast, lexical competition is considered a marker of integration into the lexicon. Qiao and Forster (2013) went on to demonstrate the prime-lexicality effect in adults only following multiple training sessions on nonwords which included meaning (Qiao & Forster, 2013). While considered evidence for competition and therefore lexicalisation, it can nonetheless be questioned whether the prime-lexicality effect itself might be driven by episodic interference, biases or confusion.

In Qiao and Forster (2013), only half of the exposed novel words appeared as related primes (e.g. *banara – BANANA*, following training on *banara*), while the other half did not appear as primes for their basewords (e.g. *forever – BATTERY*, following training on *baltery*). The authors suggested that slowed down responses to base words in the absence of novel word primes (e.g. *forever – BATTERY*) likely constitute some kind of episodic interference. In line with this, responses to basewords of exposed novel words were overall slower than responses to basewords of unexposed novel words, whether or not basewords were primed by their novel word neighbours. In other words, responses to base words were slowed, even in the absence of their neighbouring novel word prime. This could indicate some episodic confusion influencing performance on the task. Crucially, however, the extent of this overall latency was modulated by the presence vs. absence of the novel word primes. There was a significant difference between base words primed by their trained novel word neighbours relative to basewords (also neighbours of trained words) primed by unrelated primes. This showed an inhibitory priming component that is specific to the presence of the trained novel word paired as a prime with a target overlapping in form. Given the brief and masked presentation of the primes, it is unlikely that this specific inhibitory effect was driven by episodic interference. It is however consistent with lexical competition.

While the above design has the advantage of separating episodic interference and lexical competition, it was not adopted in its exact form in the present experiments. The overall aim of this thesis was to investigate incidental learning via stories and thus, novel words were embedded rich, relatively lengthy contexts. Exposing children to such materials is much more time-consuming than training adults on novel words in isolation (e.g. Bowers et al., 2005) or with short definitions (e.g. Qiao & Forster, 2013). Inevitably however, it limits the number of novel words each child can be exposed to. Furthermore, working with children in school settings and using words rather than nonwords introduces variability and noise, which is why performance was compared between pre-exposure and post-exposure, within-subjects and within-items. To maximise power, all ten novel words experienced by a given child in the exposure phase appeared as related primes. This allowed the detection of facilitatory as well as inhibitory priming effects. The disadvantage of this design is that unlike Qiao and Forster's (2013) paradigm, it does not allow a strict differentiation between lexical competition and episodic interference. It thus becomes important to ask, what is the likelihood that the prime-lexicality effects observed here were driven by lexical competition between novel words and their neighbours after exposure, rather than more general episodic interference?

Evidence in favour of a lexical competition account of the prime-lexicality effect is provided by Experiment 2. Here, the prime-lexicality effect only emerged after twelve exposures over four sessions, but not after four exposures in one session. This finding is difficult to explain solely by an episodic interference account, because lexical configuration of novel words (i.e. spelling and orthographic decision, argued to be supported by episodic memories) was just as good in the four-exposures condition as in the twelve-exposures condition. Hence, a prime-lexicality effect driven by episode interference of 'novel word memories' would be expected to emerge equally in both conditions; this, however, was not the case. As argued in Chapter 3, the twelve-exposure condition likely allowed for more successful lexicalisation of the novel words (driven by additional exposures or additional sessions) than the four-exposures condition. This is consistent with the emergence of a prime-lexicality effect in the twelve-exposures condition only, and with a lexical-competition account of the prime-lexicality effect itself.

However, like in Qiao and Forster (2013), inhibitory priming effects after exposure likely contained an element of episodic interference in addition to lexical competition. Given the magnitude of the mean differences observed in most of the present experiments (raw difference > 100 ms in all but Experiment 3.1), the effects are unlikely to be driven solely by lexical competition caused by the primes, but possibly also due to some episodic confusion when processing the basewords (e.g., 'Was that dungeon or dudgeon?'). While Qiao and Forster's (2013) reported a unique component driven by the presence of the trained prime, this has yet to be demonstrated in children. Nonetheless, it should be noted that even if the prime-lexicality effect in children was solely driven by interference by episodic memories, this would still indicate learning. For example, episodic memories are central to instance-based models of visual word recognition, which represent individual experiences as memory traces (Reichle & Perfetti, 2003), and form an indispensable part of the Complementary Learning Systems account (McClelland et al., 1995). Thus, while this thesis advocated the lexical competition account of the prime-lexicality effect, both lexical competition and episodic interference are compatible with it being a measure of learning. Across all of the present experiments, learning was indeed demonstrated, as exposure to novel words via reading experience caused a change in priming effects, supported by further evidence from orthographic decision and spelling. Having established the presence of a prime-lexicality effect in children, future research should continue to capitalising on this novel, sensitive measure in training studies, encompassing a greater range of ages and abilities, to understand the learning processes that allow children to become skilled readers.

Given the richness of everyday reading experience, there are likely many parameters still to explore, as suggested throughout this thesis.

To conclude, the experiments in this thesis demonstrated that incidental exposure to novel written words via reading experience is a powerful learning opportunity for children. Across experiments, children showed good learning of spellings and meanings in the absence of direct instruction. Using the prime-lexicality effect as a novel measure of lexical engagement in children, orthographic learning was examined beyond the self-teaching tradition. Converging evidence emerged that children are robust orthographic learners, but future research is necessary to further uncover the mechanisms underlying this ability.

References

Adelman, J. S., Brown, G. D., & Quesada, J. F. (2006). Contextual diversity, not word frequency, determines word-naming and lexical decision times. *Psychological Science*, *17*(9), 814-823.

Andrews, S., & Lo, S. (2012). Not all skilled readers have cracked the code: Individual differences in masked form priming. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *38*(1), 152.

Bakker, I., Takashima, A., van Hell, J. G., Janzen, G., & McQueen, J. M. (2014). Competition from unseen or unheard novel words: Lexical consolidation across modalities. *Journal of Memory and Language*, *73*(1), 116–130.

Balota, D.A., Cortese, M.J., & Pilotti, M. (1999). Item-level analyses of lexical decision performance: Results from a mega-study. In *Abstracts of the 40th Annual Meeting of the Psychonomics Society* (p. 44). Los Angeles, CA: Psychonomic Society.

Balota, D. A., Cortese, M. J., Sergent-Marshall, S. D., Spieler, D. H., & Yap, M. (2004). Visual word recognition of single-syllable words. *Journal of Experimental Psychology: General*, *133*(2), 283.

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, *68*(3), 255–278.

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, *67*(1), 1–48.

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.

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.

Bowey, J. A., & Muller, D. (2005). Phonological recoding and rapid orthographic learning in third-graders' silent reading: a critical test of the self-teaching hypothesis. *Journal of Experimental Child Psychology*, *92*(3), 203–19.

Brysbaert, M., & New, B. (2009). Moving beyond Kucera and Francis: a critical evaluation of current word frequency norms and the introduction of a new and improved word frequency measure for American English. *Behavior Research Methods*, *41*(4), 977–90.

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., & Nation, K. (2006). How does orthographic learning happen? In S. Andrews (Ed.), *From Inkmarks to Ideas: Current Issues in Lexical Processing*, 151–179. Hove, UK: Psychology Press.

Castles, A., & Nation, K. (2008). Learning to be a good orthographic reader. *Journal of Research in Reading*, *31*(1), 1–7.

Clark, C. (2012). Children's and Young People's Reading Today. Findings from the 2011 National Literacy Trust's annual survey. London: National Literacy Trust.

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

Cunningham, A. E., Perry, K. E., Stanovich, K. E., & Share, D. L. (2002). Orthographic learning during reading: examining the role of self-teaching. *Journal of Experimental Child Psychology*, *82*(3), 185–199.

Davis, C. J., & Lupker, S. J. (2006). Masked inhibitory priming in english: evidence for lexical inhibition. *Journal of Experimental Psychology. Human Perception and Performance*, *32*(3), 668–687.

Davis, M. H., & Gaskell, M. G. (2009). A complementary systems account of word learning: neural and behavioural evidence. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *364*(1536), 3773–800.

de Jong, P. F., & Share, D. L. (2007). Orthographic Learning During Oral and Silent Reading. *Scientific Studies of Reading*, *11*(1), 55–71.

Duff, F. J., & Hulme, C. (2012). The Role of Children's Phonological and Semantic Knowledge in Learning to Read Words. *Scientific Studies of Reading*, *16*(6), 504–525.

Ehri, L. C. (2005). Learning to read words: Theory, findings, and issues. *Scientific Studies of Reading*, *9*(2), 167–188.

Ehri, L. C. (2014). Orthographic Mapping in the Acquisition of Sight Word Reading, Spelling Memory, and Vocabulary Learning. *Scientific Studies of Reading*, *18*(1), 5–21.

Ender, A. (2014). Implicit and Explicit Cognitive Processes in Incidental Vocabulary Acquisition. *Applied Linguistics*, 1–26.

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.

Frishkoff, G. A., Collins-Thompson, K., Perfetti, C. A., & Callan, J. (2008). Measuring incremental changes in word knowledge: Experimental validation and implications for learning and assessment. *Behavior Research Methods*, *40*(4), 907–925.

Gaskell, M. G., & Dumay, N. (2003). Lexical competition and the acquisition of novel words. *Cognition*, *89*(2), 105–132.

Gough, P. B., & Tunmer, W. E. (1986). Decoding, Reading, and Reading Disability. *Remedial and Special Education*, *7*(1), 6–10.

Grainger, J., & Ziegler, J. C. (2011). A dual-route approach to orthographic processing. *Frontiers in psychology*, *2*(56), 1-13.

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.

Hayes, D. P., & Ahrens, M. G. (1988). Vocabulary simplification for children: a special case of "motherese"? *Journal of Child Language*, *15*(02), 395.

Henderson, L. M., Weighall, A. R., Brown, H., & Gaskell, M. G. (2012). Consolidation of vocabulary is associated with sleep in children. *Developmental Science, 15*(5), 674–87.

Hills, T. T., Maouene, J., Riordan, B., & Smith, L. B. (2010). The associative structure of language: Contextual diversity in early word learning. *Journal of Memory and Language*, *63*(3), 259-273.

Hoffman, P., Lambon Ralph, M. a, & Rogers, T. T. (2013). Semantic diversity: a measure of semantic ambiguity based on variability in the contextual usage of words. *Behavior Research Methods*, *45*(3), 718–30.

Hoffman, P., Rogers, T. T., & Ralph, M. A. L. (2011). Semantic diversity accounts for the "missing" word frequency effect in stroke aphasia: insights using a novel method to quantify contextual variability in meaning. *Journal of Cognitive Neuroscience*, *23*(9), 2432–46.

Hoffman, P., & Woollams, A. M. (2015). Opposing Effects of Semantic Diversity in Lexical and Semantic Relatedness Decisions. *Journal of Experimental Psychology: Human Perception and Performance, 41*(2), 385–402.

Horst, J. S., Parsons, K. L., & Bryan, N. M. (2011). Get the story straight: contextual repetition promotes word learning from storybooks. *Frontiers in Psychology*, *2*, 17.

Hulme, C., & Snowling, M. J. (2014). The interface between spoken and written language: developmental disorders. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *369*(1634).

Joe, A. (1998). What Effects Do Text-based Tasks Promoting Generation Have on Incidental Vocabulary Acquisition? *Applied Linguistics*, *19*(3), 357–377.

Johns, B. T., Dye, M., & Jones, M. N. (2015). The influence of contextual diversity on word learning. *Psychonomic Bulletin & Review*, 1-7.

Jones, M. N., Johns, B. T., & Recchia, G. (2012). The role of semantic diversity in lexical organization. *Canadian Journal of Experimental Psychology/Revue Canadienne de Psychologie Expérimentale*, *66*(2), 115–124.

Joseph, H. S. S. L., Nation, K., & Liversedge, S. P. (2013). Using eye movements to investigate word frequency effects in children's sentence reading. *School Psychology Review*, *42*(2), 207–222.

Joseph, H. S. S. L., Wonnacott, E., Forbes, P., & Nation, K. (2014). Becoming a written word: eye movements reveal order of acquisition effects following incidental exposure to new words during silent reading. *Cognition*, *133*(1), 238–48.

Keuleers, E., & Balota, D. A. (2015). Megastudies, crowdsourcing, and large datasets in psycholinguistics: An overview of recent developments. *Quarterly Journal of Experimental Psychology*, *68*(8), 1457–68.

Kouider, S., & Dehaene, S. (2007). Levels of processing during non-conscious perception: a critical review of visual masking. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *362*(1481), 857–75.

Kwok, R. K. W., & Ellis, A. W. (2015). Visual word learning in skilled readers of English. *Quarterly Journal of Experimental Psychology (2006)*, *68*(2), 326–49.

Kyte, C. S., & Johnson, C. J. (2006). The role of phonological recoding in orthographic learning. *Journal of Experimental Child Psychology*, *93*(2), 166–85.

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), 114–33.

Leach, L., & Samuel, A. G. (2007). Lexical configuration and lexical engagement: when adults learn new words. *Cognitive Psychology*, *55*(4), 306–53.

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

Martin-Chang, S., & Levesque, K. (2013). Taken out of context: differential processing in contextual and isolated word reading. *Journal of Research in Reading*, *36*(3), 330-349.

Martin-Chang, S. L., Levy, B. A., & O'Neil, S. (2007). Word acquisition, retention, and transfer: Findings from contextual and isolated word training. *Journal of Experimental Child Psychology*, *96*(1), 37–56.

McClelland, J. L., McNaughton, B. L., & O'Reilly, R. C. (1995). Why there are complementary learning systems in the hippocampus and neocortex: insights from the successes and failures of connectionist models of learning and memory. *Psychological Review*, *102*(3), 419–57.

McClelland, J. L., & Rumelhart, D. E. (1981). An interactive activation model of context effects in letter perception: I. An account of basic findings. *Psychological review, 88*(5), 375.

McDonald, S. A., & Shillcock, R. C. (2001). Rethinking the word frequency effect: The neglected role of distributional information in lexical processing. *Language and Speech*, *44*(3), 295-322.

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.

Mol, S. E., & Bus, A. G. (2011). To read or not to read: a meta-analysis of print exposure from infancy to early adulthood. *Psychological Bulletin*, *137*(2), 267–96.

Nagy, W. E., Anderson, R. C., & Herman, P. a. (1987). Learning Word Meanings From Context During Normal Reading. *American Educational Research Journal*, *24*(2), 237–270.

Nagy, W. E., Herman, P. a., & Anderson, R. C. (1985). Learning words from context. *Reading Research Quarterly*, *20*(2), 233–253.

Nation, K., Angell, P., & Castles, A. (2007). Orthographic learning via self-teaching in children learning to read English: Effects of exposure, durability, and context. *Journal of Experimental Child Psychology*, *96*(1), 71–84.

Nation, K., & Castles, A. (in press). Putting the learning into orthographic learning. *Scientific Studies of Reading.*

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.

Norris, D. (2013). Models of visual word recognition. *Trends in Cognitive Sciences*, *17*(10), 517–24.

Ouellette, G., & Fraser, J. R. (2009). What exactly is a yait anyway: The role of semantics in orthographic learning. *Journal of Experimental Child Psychology*, *104*(2), 239–251.

Perea, M., Soares, A. P., & Comesaña, M. (2013). Contextual diversity is a main determinant of word identification times in young readers. *Journal of Experimental Child Psychology*, *116*(1), 37–44.

Perfetti, C., and Hart, L. (2002). The lexical quality hypothesis. In: L. Verhoeven, C. Elbrow, & P. Reitsma (Eds.), *Precursors of Functional Literacy* (pp. 67-87). Amsterdam: John Benjamins.

Plummer, P., Perea, M., & Rayner, K. (2014). The influence of contextual diversity on eye movements in reading. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *40*(1), 275–83.

Qiao, X., & Forster, K. I. (2013). Novel word lexicalization and the prime lexicality effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *39*(4), 1064–1074.

Qiao, X., Forster, K. I., & Witzel, N. (2009). Is banara really a word? *Cognition*, *113*(2), 254–257.

R Core Team (2014). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin*, *124*(3), 372–422.

Reichle, E. D., & Perfetti, C. A. (2003). Morphology in word identification: A word-experience model that accounts for morpheme frequency effects. *Scientific Studies of Reading*, *7*(3), 219-237.

Richter, T., Isberner, M.-B., Naumann, J., & Neeb, Y. (2013). Lexical Quality and Reading Comprehension in Primary School Children. *Scientific Studies of Reading*, *17*(6), 415–434.

Ricketts, J., Davies, R., Masterson, J., Stuart, M., & Duff, F. J. (submitted). Probing the nature of the relationship between semantic knowledge and word reading: evidence from six-year-old children.

Rodd, J. M., Berriman, R., Landau, M., Lee, T., Ho, C., Gaskell, M. G., & Davis, M. H. (2012). Learning new meanings for old words: effects of semantic relatedness. *Memory & Cognition*, *40*(7), 1095–1108.

Rumelhart, D. E., & McClelland, J. L. (1982). An interactive activation model of context effects in letter perception: II. The contextual enhancement effect and some tests and extensions of the model. *Psychological review*, *89*(1), 60.

Schneider, W., Eschman, A., & Zuccolotto, A. (2007). *E-prime: User's guide, Version 2.0.* Psychology Software Tools.

Seidenberg, M. S., & McClelland, J. L. (1989). A distributed, developmental model of word recognition and naming. *Psychological review*, *96*(4), 523.

Share, D. L. (1995). Phonological recoding and self-teaching. *Cognition*, *55*(2), 151–218.

Share, D. L. (1999). Phonological recoding and orthographic learning: A direct test of the self-teaching hypothesis. *Journal of Experimental Child Psychology*, *72*(2), 95–129.

Share, D. L. (2004). Orthographic learning at a glance: On the time course and developmental onset of self-teaching. *Journal of Experimental Child Psychology*, *87*(4), 267-298.

Snowling, M. J., & Hulme, C. (2011). Evidence- based interventions for reading and language difficulties: Creating a virtuous circle. *British Journal of Educational Psychology, 81*(1), 1-23.

Sobel, H. S., Cepeda, N. J., & Kapler, I. V. (2011). Spacing effects in real- world classroom vocabulary learning. *Applied Cognitive Psychology, 25*(5), 763-767.

Swanborn, M. S. L., & de Glopper, K. (1999). Incidental Word Learning While Reading: A Meta-Analysis. *Review of Educational Research*, *69*(3), 261–285.

Taylor, J. S. H., Duff, F. J., Woollams, A. M., Monaghan, P., & Ricketts, J. (2015). How Word Meaning Influences Word Reading. *Current Directions in Psychological Science*, *24*(4), 322–328.

Taylor, J. S. H., Plunkett, K., & Nation, K. (2011). The influence of consistency, frequency, and semantics on learning to read: an artificial orthography paradigm. *Journal of Experimental Psychology Learning Memory and Cognition*, *37*(1), 60–76.

Torgesen, J. K., Wagner, R. K., & Rashotte, C. A. (1999). *Test of Word Reading Efficiency.* Austin, TX: Pro-Ed.

Tulving, E., & Thomas, D. M. (1973). Encoding specificity and retrival processes in episodic memory. *Psychological Review*, *80*(5), 352–373.

van den Boer, M., & de Jong, P. F. (2015). Parallel and serial reading processes in children’s word and nonword reading. *Journal of Educational Psychology*, *107*(1), 141–151.

van den Boer, M., de Jong, P. F., & Haentjens-van Meeteren, M. M. (2013). Modeling the Length Effect: Specifying the Relation With Visual and Phonological Correlates of Reading. *Scientific Studies of Reading*, *17*(4), 243–256.

Wang, H. C., Castles, A., Nickels, L., & Nation, K. (2011). Context effects on orthographic learning of regular and irregular words. *Journal of Experimental Child Psychology*, *109*(1), 39–57.

Wang, H.-C., Nickels, L., Nation, K., & Castles, A. (2013). Predictors of Orthographic Learning of Regular and Irregular Words. *Scientific Studies of Reading*, *17*(5), 369–384.

Wilkinson, K. S., & Houston-Price, C. (2013). Once upon a time, there was a pulchritudinous princess...: The role of word definitions and multiple story contexts in children's learning of difficult vocabulary. *Applied Psycholinguistics, 34*(03), 591-613.

Yap, M. J., Pexman, P. M., Wellsby, M., Hargreaves, I. S., & Huff, M. J. (2012). An abundance of riches: cross-task comparisons of semantic richness effects in visual word recognition. *Frontiers in Human Neuroscience, 6*, 72.

Ziegler, J. C., Perry, C., & Zorzi, M. (2014). Modelling reading development through phonological decoding and self-teaching: implications for dyslexia. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences, 369*(1634).

Appendices

Appendix 2.1

Novel word – base word pairings divided into two counterbalanced lists.

List	Novel word	Base word
A	allay	alley
	bruit	fruit
	abort	about
	envoy	enjoy
	felon	melon
	abraid	afraid
	propel	proper
	incite	invite
	pillage	village
	secrete	secrets
B	croup	group
	pivot	pilot
	antic	attic
	wield	field
	abode	above
	suture	future
	warble	marble
	palate	palace
	dudgeon	dungeon
	undermine	underline

Unrelated word – base word pairings divided into two counterbalanced lists

List	**Unrelated prime**	**Base word**
A	preen	alley
	quell	fruit
	neigh	about
	lynch	enjoy
	bridle	melon
	beadle	afraid
	entice	proper
	regale	invite
	patents	village
	commend	secrets
B	adorn	group
	wrest	pilot
	exalt	attic
	emery	field
	smite	above
	decant	future
	deride	marble
	arcane	palace
	curtail	dungeon
	zealously	underline

Nonword trials (used for both list A and B)

	Prime	**Nonword**
related	throb	throg
	twine	twide
	lance	yance
	dreary	drears
	frond	fronp
	streak	spreak
	steels	smeels
	stoics	stoins
	alloted	alloked
	areobatic	abrobatic
unrelated	liens	greft
	cameo	deace
	sprit	ghare
	carob	jench
	gruel	prese
	abject	drance
	ethers	gemes
	grange	chowds
	guilder	daughty
	factors	threves

Appendix 2.2

Stimuli for semantic relatedness task (YES trials)

The narrow road behind Freddie's house was dark and gloomy.	*alley*
Rita put lots of apples into the pie she was baking.	*fruit*
Ronnie liked to read books with lots of information on dinosaurs.	*about*
Harriet hoped that her little sister would have fun at party.	*enjoy*
After eating the juicy fruit Laura's hands were wet and sticky.	*melon*
Julie was scared that her cake might get burnt in the oven.	*afraid*
Tommy did not know the right way to eat a lobster.	*proper*
William was happy he could ask his friends over after school.	*invite*
Most of the people in the small town worked in the mines.	*village*
Helen told Kayla not to tell anyone about the stories.	*secrets*
Henry was part of a club that met every Friday.	*group*
The man flying the plane held on to the steering wheel.	*pilot*
The old wardrobe was in the room at the top of the house.	*attic*
The pitch behind the school was used for football games.	*field*
The view of the city from up high was magnificent.	*above*
When she was older Casey wanted to travel the world.	*future*
Tom gave his friend a little glass ball to cheer him up.	*marble*
The maid was cleaning the castle before the queen got up.	*palace*
On her holiday Emily visited a smelly cellar in an old prison.	*dungeon*
Jane used a red pen to highlight all the tricky words.	*underline*

NO trials

Louis was excited to meet his friends at the playground	*beans*
Daniel was proud to show his parents his maths test.	*beach*
Tina chose her favourite fairy tale as her bedtime story.	*alive*
Paula wrote a long letter to her grandmother every week.	*fence*
Katie was very happy about her new pair of shoes.	*float*
Paul did not like very soggy cereal in the morning.	*guess*
Martin was by far the fastest runner in his class.	*heavy*
Maria was excited she found a silver coin under the sofa.	*lorry*
Bradley did not get lost on the way to his new school.	*metal*
Ruth heard a strange sound coming from the garden last night.	*shell*
Jordan needed just two more stickers for his football book.	*autumn*
The music was really loud so Aaron did not hear the phone.	*change*
Chloe's favourite pudding by far was chocolate ice cream with sauce.	*nearby*
The young bride looked beautiful in her big wedding dress.	*office*
The twin brothers looked so similar it confused many people.	*pocket*
Max practised his new song on the guitar for many hours.	*feather*
After school Lucy and her two friends went blackberry picking.	*captain*
Kevin's pet rabbit ate up to ten carrots per day.	*library*
Eric blew out all the candles on the Christmas tree.	*vegetable*
Jessie did not mind watching her little sister play after school.	*bushes*

Appendix 2.3

Experiment 2 story materials

Additional texts framing the story contexts

Prologue

Doctor Dicto has stolen all the dictionaries in the world. He wants to have all the words to himself. Imagine a world without any words to say, read, write or hear! Mia and Tom are on a mission to stop Doctor Dicto's evil plan. They know he has hidden the words in lots of secret places. Luckily they found his map that shows them all the places. Each place is locked by a secret code word. Doctor Dicto used some very tricky words that hardly anyone remembers. Mia and Tom have to find out about these words to crack the codes. The only way to crack a code word is to find out what it means and enter the meaning into the code lock. Then they can free all the words that got stolen.

Epilogue

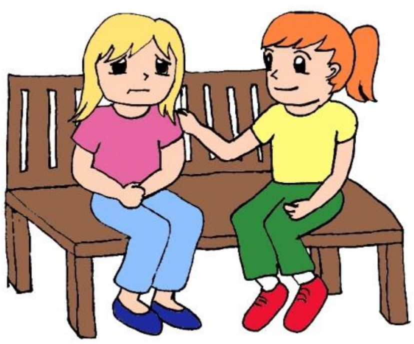
Mia and Tom have managed to find all of Doctor Dicto's hiding places. All the words are free again and can be put back into dictionaries. Mia and Tom are very happy there will still be words around to read in books, write in stories, and sing in songs.

Story contexts

List A

The forest

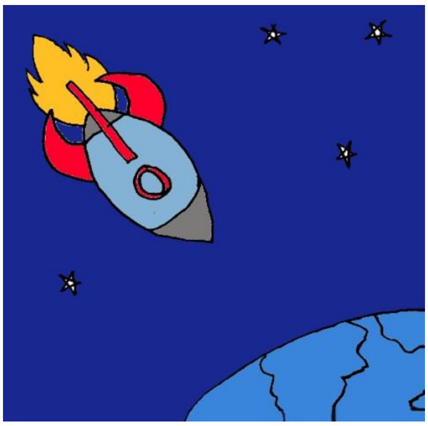
The next secret place is in a big forest. Doctor Dicto has locked the words in a hut. They need to crack the code word allay. "Let's call our teacher Miss Phips," says Tom. "To allay someone's worries is to make someone feel better and less scared," she says. "That is like when I was scared of going to the doctor," Mia says. "Mum told me he would only look in my throat." "Yes, your mum tried to allay your fears about the doctor. It is a nice thing to allay someone's fears," Miss Phips says. "I will try to allay my friends' fears in the future," Tom says. Now they know what allay means, Tom and Mia can free the words from the hut!



The wooden chest

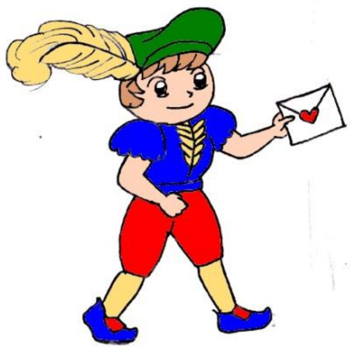
Mia and Tom find the wooden chest where the next bunch of words is locked. This time the code word is bruit. "Let’s call Auntie Jo," Tom says. "To bruit means to talk about something to lots of people," Auntie Jo says. "Often people who bruit spread a rumour or gossip." "This happened last week at school," Tom says. "Our teacher wore her slippers to school instead of proper shoes. A girl called Suzie told all the other children about it." "Yes, Suzie tried to bruit about your teacher’s slippers," Auntie Jo says. "I don’t really like it when people bruit. You can hurt people's feeling," Mia says. "But I am happy we found out what bruit means so we can crack the code!"

The tree house



Mia and Tom find the next secret place in a tree house. The door is locked. The code word to open the door is abort. Mia and Tom don’t know what abort means. They decide to call their older sister Helen. Helen thinks about how best to explain the word. "Usually it is pilots on planes who abort their flight," she says. "To abort a flight is to stop it. For example, a space shuttle might abort its flight because of a problem and go back to earth." "I think this happened when we went on holiday once. The plane had to abort its flight and go back to the airport," Tom says. Helen got the code word right. Mia and Tom can now free the words from the tree house!

The treasure chest

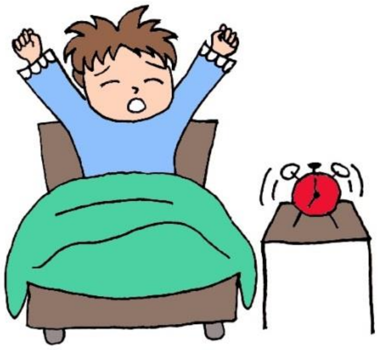


The next secret hiding place is under a bridge. The stolen words are locked in a treasure chest. The code word to free them is envoy. Their brother Mikey can help them this time. "An envoy is a person working as a messenger," he says. "The king might send out an envoy to tell his people that he is holding a ball. Then the envoy goes from house to house to tell the message. An envoy can bring both good and bad news. But these days you don’t really need to send out an envoy to deliver your news, you can just use the phone." Mia and Tom try to open the treasure chest now they know what the code word means. The chest opens and the words are free!

The piggy bank

The next set of words is locked in a piggy bank hidden in a playroom. The code word to free the words is felon. "I think I heard the word felon on TV," Tom says. Mia and Tom send an email to their cousin Annabel in America. She writes back. "Felon is a word that Americans use for some criminals. Not every criminal is called a felon. Normally it means you have done something really bad. A person is called a felon if they have to go to prison for longer than a year." "A year is a very long time to spend in prison," Mia says. "I don't want to be a felon." Luckily Annabel is right about the code word. Mia and Tom can free the words from the piggy bank.

The barn



The children arrive at the next secret hiding place. It is an old barn full of hay. The code word to open the barn is abraid. Mia and Tom call their brother Markus for help. "Abraid is a word people used in the olden days," Markus explains. "Abraid simply means to wake up in the morning." "Does anyone still use the word abraid?" Tom asks. "I have never heard it." "Good question," says Markus. "Maybe old people still do. Maybe when the Queen has to get up in the morning, her butler tells her that it is time to abraid." "Tomorrow morning I will tell Tom to abraid and get ready for school," Mia says. Mia and Tom quickly free the words from the barn.

The bottle

The next bunch of words is locked in a bottle floating in a pond. Mia and Tom get the bottle. They have to find out what the code word propel means. Next to the pond is an old lady walking her dog. Tom asks her if she can help them with the word propel. "To propel something is to move it forward quickly," says the old lady. "For example, when you sit in a boat you use a paddle to propel yourself through the water." "I once used a shovel to propel myself forward in a boat," says Tom. "Great idea," says the lady. "Some big boats have a motor or an engine that helps to propel them." Mia and Tom can now open the bottle.

The secret garden

Tom and Mia next go to a secret garden. The words are locked in a small shed. The code word is incite. Tom calls his history teacher. "I can help you with the word incite," Mr Magpie says. "To incite someone is to encourage them to do something bad. I could tell Mia to go and hide the teacher’s notebook. I could offer her my pudding at lunch time to incite her to do it. So I am trying to incite Mia to be naughty by offering her a treat." "I get it," says Tom. "Last night Bradley tried to incite me to put a worm in Molly’s bag. But I didn’t do it." Mia and Tom can now free the words from the shed!

The trap door

Mia and Tom go to the next secret place. It’s under a trap door in an old house. The trap door is locked by the word pillage. This time Tom has an idea. "I think I heard about the word pillage when I was reading about the Vikings", he says. Tom gets out his history book and reads. "A thousand years ago the Vikings used to pillage many houses. To go around and pillage means to steal from a place. The Vikings broke into houses and stole things like food and money." "That sounds like what burglars might do. To pillage a house is not very nice," Mia says. But the word pillage will help Mia and Tom free lots of words. Now they can open the trap door!

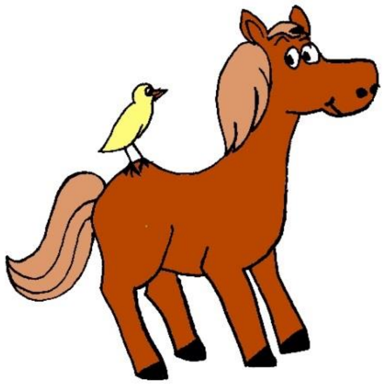
The suit of armour

Mia and Tom find the next hiding place in a museum. The words are locked in a knight’s suit of armour. To free the words, they have to find out about the word secrete. Mia and Tom ask the museum guard for help. "Secrete is another word for hiding something," the guard explains. "For example, Doctor Dicto’s plan is to secrete away all the words in the world. Did you ever secrete something?" "Once I hid lots of chocolate bars under my bed, but Mum found them," Tom says. "Yes, you tried to secrete the chocolate, but it didn't quite work," says the guard. "We must free the words now," Mia says. "Thank you for all your help!"

List B

The wardrobe

The word Mia and Tom have to crack is croup. This will unlock lots of words trapped in an old wardrobe. "I think I read about croup in a book about horses and ponies," Mia says. Mia and Tom walk to a stable with horses down the road. They find a man brushing a horse. "Excuse me. Do you know what the word croup means?" Mia asks. Luckily the man can help them. "The croup is the highest part of its back. If you imagine sitting on a horse, its croup will be behind you. At the end of the croup is the horse’s tail," he explains. Mia and Tom are very happy. They can go back to the wardrobe and free the words!



The suitcase

Mia and Tom find the next words locked in an old suitcase. The code word to open the suitcase is pivot. Mia and Tom decide to ask their neighbour Miss Doodle. Luckily Miss Doodle can explain to them what pivot means. "To pivot means to spin around on the spot. For example, a spinning top can pivot when you twist it," she says. "You can also pivot around on your own heel until you get dizzy." "In dance class we sometime spin around on one foot," Mia says. "It is really tricky not to lose your balance." "Yes, that is another way to pivot," says Miss Doodle. Mia and Tom now know what the code word means. They can open the suitcase and free the words.

The phone box

The next words are locked in an old phone box. The code word to open the phone box and free the words is antic. Mia and Tom ask a policeman on the street. "Antic is another word for a joke or a prank you can play on someone," the man says. "Do you get into trouble for an antic?" asks Mia. "Not necessarily," says the policeman. "An antic can simply be a funny action that makes people laugh. For example, walking around town in a clown costume is an antic. But it wouldn’t get you into trouble." "Thank you for telling us what antic means," says Tom. Mia and Tom are happy they can free the words from the phone box.

The church tower

Mia and Tom find the next bunch of words locked in a church tower. They have to find out what the code word wield means. The priest working at the church can help them. "To wield a weapon means to hold it out ready to attack," he explains. "A knight might draw his sword and wield it at a dragon. This means the knight is ready to defend himself. Can you think of another example?" "A cowboy might wield a pistol at a thief trying to steal his cows." says Tom. "A wizard might wield his wand to cast a spell," says Mia. "I see you understand what wield means," says the priest. Now Mia and Tom can free the words from the tower!

The old abandoned house

Mia and Tom find the next hiding place in the ruins of an old abandoned house. The door to the room with the words is locked. The code word is abode. They decide to call Uncle Jimmy. "I know what abode means," Jimmy says. "It is a word that fancy people in old books use. Abode means the place where someone lives. 'This is my abode' means the same thing as 'this is my home'." "So this house where the words are hidden used to be someone’s abode," says Tom. "Yes," says Uncle Jimmy. "When Doctor Dicto picked abode as the code word for this place, he gave you a clue!" Tom and Mia can now free the words from the room in the old abandoned house.

The grandfather clock

Mia and Tom manage to find the next hiding place. The stolen words are locked in an old grandfather clock. The code word to free the words is suture. They ask their Dad if he knows what it means. "To suture a cut means to stitch it up. A doctor can do that if a plaster is not enough," Dad tells them. "For example, once I fell down and cut my hand on a branch. The doctor had to suture the cut so it could heal." "Does it hurt to suture a cut?" asks Mia. "Not really," says Dad. "Before the doctor starts to suture your hand, he makes it go numb so you can’t feel anything." Mia and Tom have cracked the code word suture and can free the words!

The cave

Mia and Tom get to a cave. The entrance is locked by a big metal gate. The gate is locked by the code word warble. They call their granny. "Hmm," says granny. "I am not sure about warble. I know warble is a type of singing. I will ask my friend Lucy. She is a singer at a theatre." Thank goodness Lucy can tell them what warble means. "To warble means to sing in a funny way like a chirpy bird," Lucy shows them. "Ahahahah," she sings. Warble sounds a bit like wobble, Mia thinks. Lucy’s voice is wobbling up and down. Mia and Tom type the answer into the code lock. The gate opens. Now the words are free.

The Ferris wheel

The next hiding place is on a Ferris wheel at the fun fair. The code word to open the box with the words is palate. Mia and Tom ask the man selling tickets for help with the code word palate. "Your palate is your sense of taste when it comes to food you like or don't like. Usually, people talk about their palate when they go to a fancy restaurant. If they don’t like something, they might say that the food does not suit their palate," says the man. "Our school dinners don’t always suit my palate," says Mia. "Especially when they serve us mushy broccoli. I really don't like broccoli." "I don't like broccoli either," Tom agrees. They quickly free the words from the Ferris wheel.

The vending machine

Doctor Dicto’s next hiding place is in a vending machine full of sweets. The code word to open the vending machine is dudgeon. Mia and Tom call their Nanny Sue and tell her about their mission. She can help them. "Dudgeon is a fancy word for anger or fury. A person is in high dudgeon if they are really angry. When was the last time you saw someone in high dudgeon?" Nanny Sue asks. "Our mum was in high dudgeon when the dog ate her new shoes. She was shouting at the dog. But luckily she calmed down quickly," Tom says. "I would be angry too," Nanny Sue laughs." Thank you Nanny for telling us what dudgeon means!" Mia cheers.

The piano

The next secret hiding place is in a piano at their school. The words are locked in the piano by the code word undermine. Mia and Tom ask their music teacher Mr Flute for help. "To undermine something means to build a tunnel underneath it," says Mr Flute. "For example, a prisoner trying to escape might undermine his cell. He would have to do a lot of digging. Also, when the London underground was built, the workers had to undermine the streets to build the tunnels for the trains." "Lots of moles undermine our garden with their tunnels. Dad doesn't like it," says Mia. Mia and Tom are happy they now know what undermine means. They can free the words from the piano!

Notes. Stories were laid out on A5 cards. Story contexts shown contain six exposures per novel words. For the reading-plus condition, versions with five exposures were created. The dictionary-style definition was always preserved.

Appendix 2.4

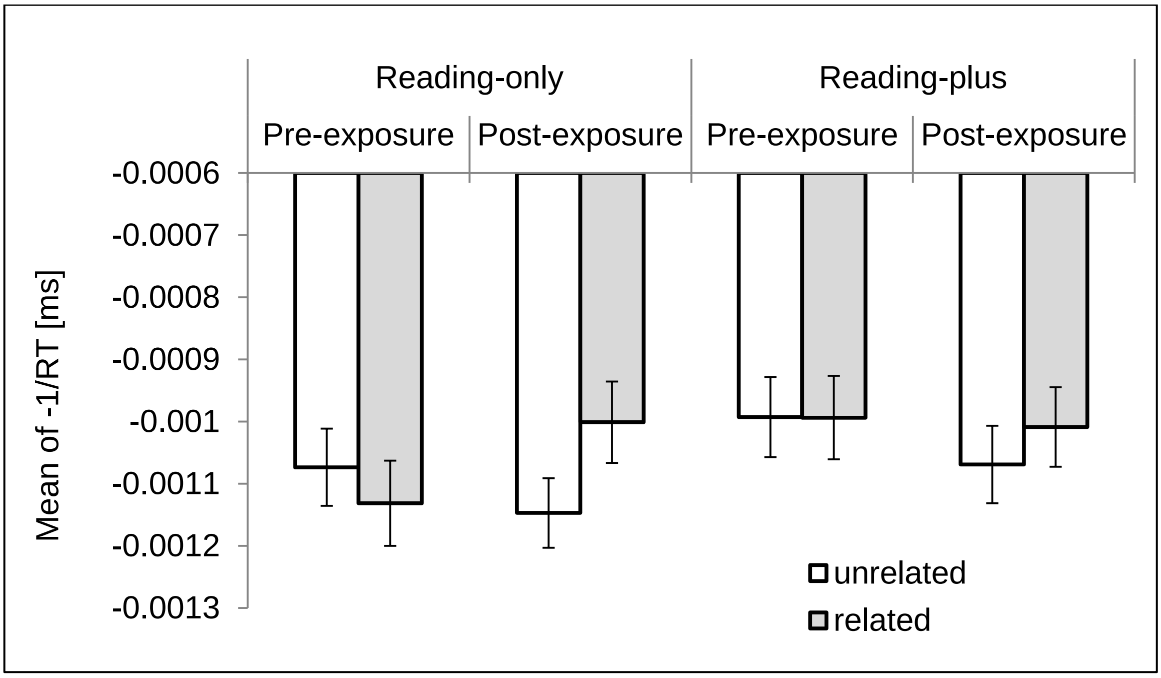
Orthographic decision items

List	correct form	homophone foil	one-letter different foil	one-letter-different foil
A	allay	aleigh	allat	allab
	bruit	broot	bruim	bruip
	abort	abawt	abork	aborp
	envoy	ennvoy	envol	envok
	felon	fellun	felop	felob
	abraid	abrayed	abrain	abraif
	propel	propell	propen	propes
	incite	insite	incife	incike
	pillage	pillege	pillane	pillaye
	secrete	sucreet	secreme	secrepe
B	croup	croop	croug	croum
	pivot	pivet	pivom	pivor
	antic	antick	antid	antiv
	wield	whield	wielk	wielp
	abode	aboad	abole	abope
	warble	worble	wardle	warkle
	suture	soucher	sutune	sutuke
	palate	palete	palape	palave
	dudgeon	dudgen	dudgeos	dudgeor
	undermine	undamine	undermike	undermite

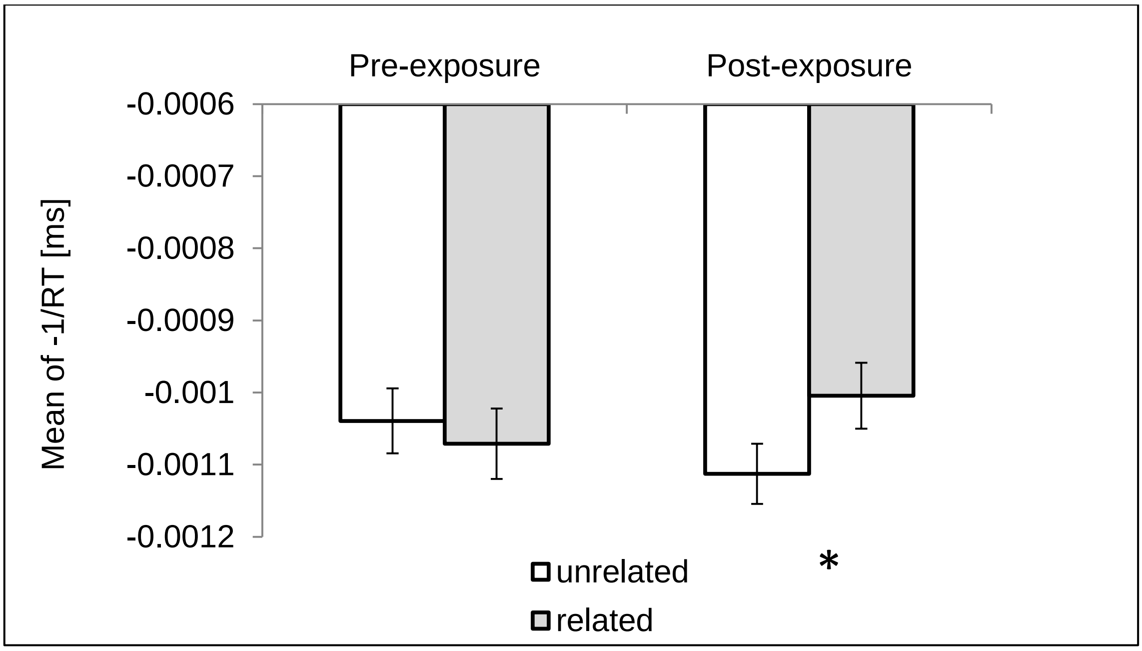
Appendix 2.5

Inverse transformed data for masked priming lexical decision

2.5.1 Data split by exposure condition. Error bars show 95% confidence intervals.



2.5.2 Data collapsed across exposure conditions. Error bars show 95% confidence intervals.



Appendix 2.6

Notes. 'Condition' refers to reading-only vs. reading-plus.

2.6.1 Orthographic decision

Formula: Accuracy ~ TOWRE + Condition + (1 | Subject) + (Condition | Item) (Family = binomial)

Fixed effects:

	Estimate	***SE***	***z***	***p***
Intercept	-4.26	2.56	-1.66	.096
TOWRE	-0.05	0.02	2.67	.023
Condition	0.43	0.66	0.65	.513

2.6.2 Spelling

Formula: Accuracy ~ TOWRE + Condition + (1 | Subject) + (Condition | Item) (Family = binomial)

Fixed effects:

	Estimate	***SE***	***z***	***p***
Intercept	-2.08	2.68	-0.78	.437
TOWRE	-0.04	0.02	1.81	.071
Condition	1.82	0.90	2.02	.043

2.6.3 Definition knowledge

Formula: Accuracy ~ TOWRE + Condition + (1 | Subject) + (Condition | Item) (Family = binomial)

	Estimate	***SE***	***z***	***p***
Intercept	-0.29	1.56	-0.86	.855
TOWRE	-0.01	0.01	1.08	.282
Condition	0.43	0.41	1.05	.294

2.6.4 Masked priming lexical decision task

Notes. 'Relationship' refers to prime-target relationship (unrelated vs. related). 'Session' refers to test-session (pre-exposure vs. post-exposure).

2.6.4.1 Main model

Formula: (-1/RT) ~ TOWRE + Condition * Session * Relationship + (Session * Relationship | Subject) + (Condition * Session * Relationship | Item)

Fixed effects:

	Estimate *(10^{-6})*	***SE*** *(10^{-6})*	***t***
Intercept	20.3	319	0.06
TOWRE	-9.61	2.82	-3.40
Condition	-3.30	86.0	-0.04
Session	-16.8	24.3	< 0.01
Relationship	46.6	21.7	2.15
Relationship x Session	146	37.4	3.91
Relationship x Condition	-26.3	37.3	-0.71
Session x Condition	-32.0	48.3	-0.66
Condition x Session x Relationship	-115	72.0	-1.60

2.6.4.2 Pairwise comparisons to explore prime-target relationship x test session interaction

Pre-exposure only:

Formula: (-1/RT) ~ TOWRE + Relationship + (Relationship | Subject) + (Relationship | item)

Fixed effects:

	Estimate *(10^{-6})*	***SE*** *(10^{-6})*	***t***
Intercept	-122	326	-0.37
TOWRE	-8.29	2.89	-2.67
Relationship	-30.1	27.0	-1.11

Post-exposure only:

Formula: (-1/RT) ~ TOWRE + Relationship + (Relationship | Subject) + (Relationship | item)

Fixed effects:

	Estimate *(10^{-6})*	***SE*** *(10^{-6})*	***t***
Intercept	-18.9	352	-0.05
TOWRE	-9.26	3.12	-2.97
Relationship	115	28.5	4.03

2.6.5 Semantic Relatedness

Notes. The maximal model did not converge; to simplify the model, the by-item random slope for session was removed. The difference in explained variance between the maximal model and the simplified final model was not significant.

Formula: log(RT) ~ TOWRE + Condition * Session * Base word status + (Session * Base word status | Subject) + (Condition * Base word status | item)

	Estimate ***(10^{-3})***	***SE*** *(10^{-3})*	***t***
Intercept	8243.94	457.76	18.01
TOWRE	-9.39	4.05	-2.32
Condition	120.92	122.74	0.99
Session	-291.07	26.91	-10.82
Base word status	-28.96	33.08	-0.88
Condition x Base word status	24.86	73.15	0.34
Session x Base word status	43.12	51.74	0.83
Condition x Session	12.57	54.12	0.23
Condition x Session x Base word status	199.54	103.89	1.92

Appendix 3.1

Story materials for Experiment 2 - 4

Additional information sheet introducing the story contexts

Story contexts

List A

1. Help for Sara

Today Sara was not looking happy. Billy asked her what was wrong. “My parents told me that we are going to visit my aunt and uncle in Sweden,” she explained. “That sounds fun,” Rachel suggested. “Don’t you like them?” “I do. My cousins are very funny,” Sara replied. “Then why aren’t you happy?” Jimmy asked. “I am scared of flying,” Sara admitted. “We must try to allay Sara’s fear of flying,” Rachel insisted. The others agreed to try to allay her fears. “Why are you scared?” Billy asked. “I have never flown before. I don’t know what it will be like,” Sara explained. “I’ve flown before,” Jimmy said, “I can tell you about it. Maybe that will allay your fears.” “Me too!” Rachel chimed in. Rachel and Jimmy told Sara all about their plane journeys to allay her fears. Sara looked much happier. “I am glad I talked to you about it. I am less scared now,” she said.

2. Nasty gossip

Today the Caterpillar Club met after school. Rachel looked worried. “There was some gossip going around today that the Caterpillar Club was over,” she told the rest of the club. “What do you mean?” Billy asked. “I heard them bruit about us, saying we were no longer a club because we had a fight,” Rachel explained. “That’s not true!” Billy shouted. “Of course it's not true,” Rachel replied. “But everyone was talking about it.” “I wonder who started to bruit about this gossip,” Sara said. “Only the Gloopy Goo Gang would bruit about us to make us look bad,” Jimmy said angrily. “So what shall we do now?” Sara asked. “We have to show everyone that we are not over!” Billy said. “How do we do that?” Jimmy asked. “We do the same we always do. We stick together. Who cares what people bruit about!” Rachel said firmly.

3. A gloopy surprise

Today the Caterpillar Club were on a mission to sneak into the Gloopy Goo Gang’s headquarters to find out about their plans. The headquarters were in a tree house in a front garden. Billy was sent ahead as a spy while the rest of the club hid behind some bushes. Billy snuck up to the tree house. There was no one in sight. Suddenly, a pair of arms holding a bucket appeared through the window of the tree house. “Watch out!” Rachel shouted. Billy looked up and was hit by a wave of green gloopy goo from the bucket. He was shocked for a moment, but then ran back to rest of the gang. “Let's abort the mission!” Billy called to them. “We can't abort, we have to keep trying!” Jimmy objected. “Whether you abort the plan or not, I need a shower,” said Billy, wiping the goo of his face. “This is quite tasty actually. It’s jelly!” So the Caterpillar Club agreed to abort the mission and plan their revenge.

4. Surprise for Mary

Sara was late arriving at the Caterpillar headquarters. "I have some bad news," she said. "Our friend Mary had a bike accident. She didn't break any bones, but she has to stay at home. She can’t come to the school party tomorrow." "Poor her," Rachel said. "She was really looking forward to the party." "I wish we could do something to cheer her up," Billy wondered. "How about we bring the party to her house! Today!" Jimmy suggested. "Great idea," Rachel cheered. "But we need help to organise a party quickly." "Someone should go as an envoy to all our friends' houses to ask them to help," Jimmy said. So Billy was sent off as an envoy while the others started preparing the surprise party. Finally the envoy returned with lots of children who wanted to help. Sara had gone as an envoy to Mary's house to ask her parents' permission. Luckily they thought it was a lovely idea.

5. Cousin Annabel

Today the Caterpillar Club were having an interesting conversation in Jimmy's back yard. "I got a card from my cousin Annabel today," Jimmy said and showed it to the others. "She lives in America. I wish I could visit her. I think she has a very interesting life." "What is she doing in America?" Rachel asked. "She works in a prison," Jimmy explained. Everyone was impressed. "Does that mean she has met a real felon who is in prison?" Billy asked. "Of course she has met a felon," Jimmy said. "She works with criminals all the time." "I wonder what it is like to meet a felon," Sara said. "I think I would be scared." "Me too," Rachel said. “A felon must have done something really bad." “Annabel isn’t scared of anything,” Jimmy explained. “Just like me!” Billy shouted. “Except spiders,” Sara said and grinned. “Yes, except spiders,” Billy admitted.

6. Revenge in the morning

The Caterpillar Club was camping in Sara's garden. The boys and girls had separate tents. During the night, the boys tried to scare the girls by making freaky noises. At first the girls were scared, but then they realised it was just the boys. "Let's go tell the boys that it is time to abraid," Sara said the next morning. They looked inside the boys' tent. They were fast asleep. "Hey boys, time to abraid!" Sara called, but the boys did not stir. "No wonder the boys won't abraid. They were up all night trying to scare us," Rachel said. "I have an idea," Sara said and grinned. "This will make sure they abraid in no time at all!" Sara got the hosepipe. She started spraying water at Jimmy and Billy's faces. The boys woke up. Rachel and Sara were laughing. "This is our revenge for last night!" Sara laughed. "Fine, we're even," Jimmy said.

7. No paddle!

Today the Caterpillar Club decided to go on a trip to a nearby lake. It was a small lake and there was no one else in sight. “Look!” exclaimed Sara. “There is a rowboat!” An old rowboat was tied to a tree at the shore. It looked like it had not been used for a very long time. “There are no paddles. What are we going to use to propel the boat forward?” Rachel asked. “Let’s look around for something useful to propel us,” Jimmy suggested. “How about this?” Billy said. He was holding up an umbrella. “I don’t think this will work to propel us,” Jimmy replied. Finally, Rachel discovered a narrow wooden plank about as long as her arm. “Let’s try using this to propel us forward!” Sara urged. They all climbed into the row boat and started rowing with the plank. The boat was moving very slowly in a zig zag line, but it worked!

8. Battle for the playground

When the Caterpillar Club arrived at the playground on this sunny day, the Gloopy Goo Gang was already there. “We were here first!” Harry shouted. “How about we battle for the playground?” Jimmy replied. He was trying to incite a battle. “The winners get to have the playground to themselves for a week,” Sara suggested. This was enough to incite the Gloopy Goo Gang to agree to a battle. “Let’s have a race around the playground,” Billy suggested. “Both teams pick their fastest runner and they will race against each other.” Jimmy was chosen to race against Harry. Billy blew his whistle. It was a close race, but in the end Jimmy finished first. “I am glad we managed to incite them to do this!” Rachel cheered. “I wish we had not let them incite us to battle for the playground,” Harry said to his gang. “We were here first after all!”

9. How to be a Viking

Today the Caterpillar Club were working on a school project. They had to give a presentation to their class about the Vikings. They were at the library looking through books. “There are too many different kings and battles in here,” Billy complained. “How about we dress up as Vikings and act out some of the things they are famous for,” Billy suggested. “It will be less boring.” “What would we do?” Rachel asked. “We could storm the classroom and try to take over, like the Vikings tried to take over England," Jimmy offered. "The Vikings used to pillage houses in order to get food and money," Sara read out. "How about we pretend to pillage?" "I suppose we could try to pillage everyone's lunch boxes," Billy suggested. So the Caterpillar Club continued to plan their special presentation. It was huge success when they finally came into school with their costumes and acted out how the Vikings used to pillage houses.

10. A box of treasures

Today Rachel brought a shoe box to the Caterpillar headquarters. "I saw something in a movie," she explained. "A gang of children decided to secrete away a box of treasures. They buried it in a forest and made a pact to come back in ten years and dig it back out." Everyone liked that idea. "I am not sure what treasure I would like to secrete away," Sara wondered. "Let's all go home and look for something to secrete away in the box," Jimmy suggested. "We can meet at the edge of the forest." So an hour later the Caterpillars met to secrete away their treasures - an old action figure from Billy, a photo from Rachel, a special coin from Jimmy and a pretty stone from Sara. They put everything in the box and buried it under a tree. Then they carved their initials into the tree to mark the spot.

List B

1. Copper's favourite

Today the Caterpillar Club was visiting the horse owned by Rachel's mum. It was called Copper, because of its light brown fur. "I don't know much about horses" Billy said. "Maybe your mum can teach us something." Rachel's mum showed them how to brush the horse. The horse was very friendly and stood calmly. "Brush along the fur with soft strokes," Rachel's mum showed them. "Does he like being brushed?" Jimmy asked. "Oh yes," Rachel's mum replied. "His favourite part is when you brush his croup." She passed the brush to Jimmy and showed him where on his back he could find the croup. It was quite close to Copper's tail. When Jimmy started to brush the croup, Copper looked very content. The Caterpillar Club all had a go at brushing Copper, paying particular attention to his croup. At the end, Copper's fur looked very clean and shiny.

2. Silly Billy

Today the Caterpillar Club were meeting at Rachel's house to work on a school assignment. She had a computer in her room. They had to find out about French painters on the internet. Billy was not in the mood to do work. "I want to play a game!" he said. "Me too," Jimmy said, "but we have to do this assignment by tomorrow." Billy started to pivot around in Rachel's desk chair. He was spinning quickly as the chair started to pivot faster. "You are making me dizzy just by watching you pivot like that!" Sara complained. "Don't look then!" Billy called. "Let's see if we can make him dizzy, too," Jimmy said. Together they started to pivot Billy around as fast as they could. Billy was laughing. "Stop, stop, I'm getting dizzy!" he finally yelled. "Let's get to work or we will never finish," Rachel urged. The Caterpillar Club found out many things about French painters that day.

3. Gloopy revenge

Today the Caterpillar Club was sitting in their headquarters. They were talking about their enemies, the Gloopy Goo Gang. “It’s time we played another prank on the Gloopy Goo Gang,” Jimmy announced. “What antic shall we get up to this time?” Sara asked. “I think we should get them with their own weapon,” Rachel suggested. “An antic with gloopy goo!” The Caterpillar Gang went to Sara’s house and cooked a big pot of green, gloopy jelly. Then they carried it to the Gloopy Goo headquarters. “What antic shall we do with it?” Billy asked. “Look,” Jimmy called. “One of them left their wellies at the bottom of their tree house! We should fill them with jelly!" So the Caterpillar Club snuck up to the wellies and performed their antic. Then they hid behind some bushes to wait for someone to come. Harry the boss finally came and put them on. He wailed loudly when he realised his foot was stuck in jelly.

4. Water fight

Today the Caterpillar Club found a note at Jimmy’s front door. It said ‘Water fight at the playground. The Gloopy Goo Gang.’ “Shall we go?” Jimmy asked. “Of course!” everyone cheered. The Caterpillar Club got together all their water guns and filled balloons with water to make water bombs. Then they went to meet the Gloopy Goo Gang. “Look who showed up to get wet!” Harry the boss yelled. He had chosen to wield a water bomb. “On three!” Jimmy shouted and took out his water gun to wield it at their enemies. “One, two, … three!” Water bombs were flying and water was shooting from the water guns. Rachel saw that Ollie picked up water bomb to wield it at her. She dodged it. It fell onto the grass and did not burst. She picked it up to wield it threateningly at the Gloopy Goo Gang. She threw it and hit Ollie in the legs. At the end of the fight, everyone was very wet.

5. Uncle Henry

Today the Caterpillar Club were visiting Jimmy's Uncle Henry. He lived in a big, old mansion. "Welcome to my abode!" Uncle Henry greeted them at the door. The Caterpillar Club stepped into the large entrance hall. There were big paintings on the wall and lots of old furniture. "Let me give you a tour of my abode," Uncle Henry offered. "This house has been my family's abode for over 300 years." There were many interesting things to see around this mansion. The portraits on the walls all showed members of Uncle Henry's family. There was even a knight's suit of armour standing in a hallway. "My great-great-great Uncle Edward wore this," Uncle Henry explained. "Can I try it on?" Billy asked. "I am afraid it might be a bit too big for you," Uncle Henry replied. Billy looked disappointed. "I hope you enjoyed your visit to my abode!" Uncle Henry said at the end.

6. Brave Jimmy

Today Jimmy arrived at the playground with a bandage around his hand. The others gathered around and looked concerned. "What happened?" Rachel asked. "I fell off my skateboard. I tripped over a rock on the street and cut my hand," Jimmy explained. "My mum had to take me to the hospital." "What did the doctors do?" Billy asked. "The doctor had to suture the cut on my hand. He said he had to suture it so it could heal properly. Then he put the bandage on." "Did it hurt to get the stitches?" Sara asked.

"Not really," Jimmy replied. "Before the doctor started to suture my hand, he made it go numb. Then I couldn't feel anything!" "You are very brave," Rachel said. "I would be quite scared if I had to let someone suture my hand." "I was a bit scared too, but then I realised that it wasn't that bad," Jimmy replied.

7. Aunt Dorie

It was just another Sunday at the Caterpillar Club headquarters. "My Aunt Dorie has invited us over for tea and cake," Rachel announced. "She makes very good cake." So the Caterpillar Club set off to Aunt Dorie's house. Aunt Dorie was an interesting looking person. She had pitch-black curly hair that she wore piled up in a huge tower twice the size of her head. "Did Rachel tell you that I was an opera singer?" She asked while pouring the tea. The Caterpillar Club shook their heads. At once she broke into a song in a foreign language. Her voice sounded high-pitched and quivering. "How do you sing like that?" Sara asked. "You have to let your voice warble," Aunt Dorie explained. "Try! Copy how I warble. This is a famous Italian song: O sole mio!" "Oh silly meow!" Billy sang in a funny voice. Everyone laughed at his attempt to warble. "Good job trying to warble!" Aunt Dorie cheered.

8. A special cake

Today the Caterpillar Club was talking about food while they sat in Rachel's kitchen. "What's the worst food you have ever eaten?" Rachel asked. "I tried snails on holiday in France once," Jimmy replied. "But they did not suit my palate." "I once had really spicy chilli," said Sara. "Spicy food rarely suits my palate." "This gives me an idea," Billy said and grinned. "How about we make a special prank cake for the Gloopy Goo Gang? We make it look like a yummy cake, but really it is made out of disgusting food." "That is a funny idea!" Sara chimed in. "We could hide some sardines in it. "Instead of chocolate sauce we could put barbecue sauce on it!" Rachel exclaimed. "I am sure that will not suit their palate." "And we could use mayonaise instead of cream!" Jimmy added. The Caterpillar Club continued to plan their magnificant prank cake that would surprise anyone's palate.

9. A nasty message

Today Jimmy called all the members of the Caterpillar Club to meet at his house. He was in high dudgeon. Someone had written a sentence with chalk in large letters on the pavement. It said: 'Caterpillars are slow and ugly!' "I bet the Gloopy Goo Gang wrote that!" Jimmy exclaimed. "What a nasty thing to do," Rachel said. "What shall we do?" "I'm in such high dudgeon about this!" Billy shouted. "And only revenge will calm my dudgeon!" "If you want revenge, write them a message too!" Sara suggested. "What would we write?" Jimmy asked. "Gloopy goo is.... gloopy?" Everyone laughed, except Billy, who was still in high dudgeon. "Let's draw butterflies all over their pavement, to show them that caterpillars can transform into anything!" Rachel suggested. "That's such a girly thing to do," Billy said grumpily. "While you try to think of something better, I'll go with Rachel," Sara said.

10. Annabel's tale

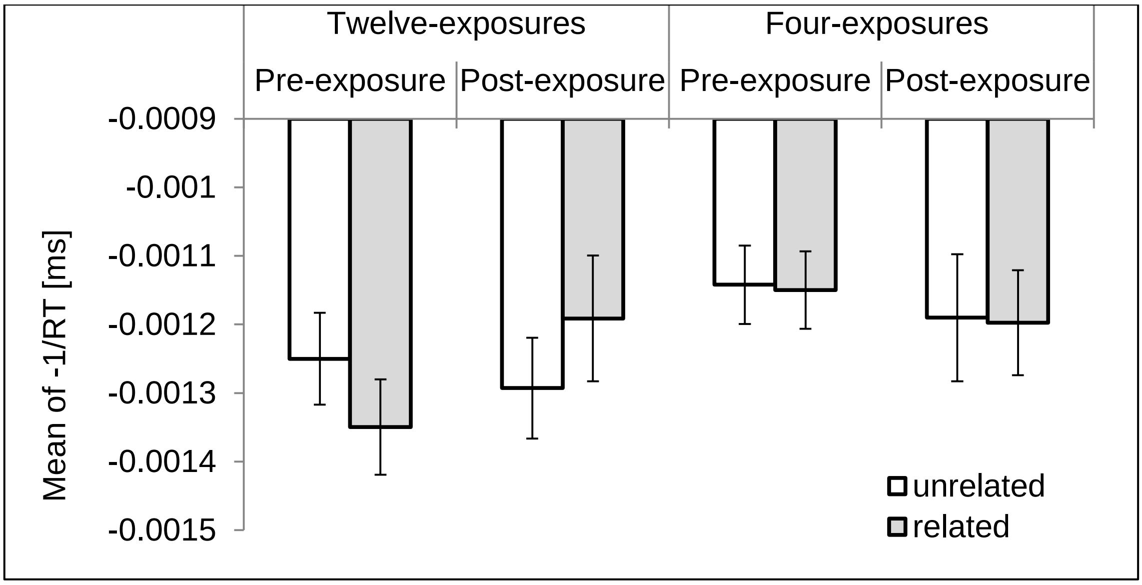
Today the Caterpillar Club was meeting at Jimmy's house. It was raining. “I got an email from my cousin Annabel today,” Jimmy said. "She works as a prison guard in America. She had a very exciting story to tell." Everyone was curious to hear about that. "Some of their prisoners tried to escape," Jimmy told them. "The tried to undermine their cell to build a tunnel." "Did they manage to escape?" Billy asked. "No," Jimmy said, "But they got quite far. Apparently they managed to undermine all the way to the edge of the prison." "I am surprised you can undermine a prison. There must be a lot of concrete in the floor," Sara wondered. "Apparently they made their own tools to undermine the floor," Jimmy explained. "But the prison guards caught them in the end. Annabel was one of them!" "That's good," Rachel said. "Those prisoners must be in there for a reason!"

Notes. Stories were laid out on A5 cards.

Appendix 3.2

Inverse transformed data for masked priming lexical decision

3.2.1 Data split by exposure condition. Error bars show 95% confidence intervals.



Appendix 3.3

Notes. 'Condition' refers to four-exposures vs. twelve-exposures.

3.3.1 Orthographic decision

Formula: Accuracy ~ Condition + (1 | Subject) + (Condition | Item) (Family = binomial)

Fixed effects:

	Estimate	***SE***	***z***	***p***
Intercept	2.59	0.54	4.84	<.001
Condition	-1.16	0.95	-1.22	.224

3.3.2 Spelling

Formula: Accuracy ~ TOWRE + Condition + (1 | Subject) + (Condition | Item) (Family = binomial)

Fixed effects:

	Estimate	***SE***	***z***	***p***
Intercept	-3.19	2.75	-1.16	.246
TOWRE	-0.05	0.02	1.96	.0496
Condition	-0.2	0.66	-0.30	.762

3.3.3 Definition knowledge

Notes. The maximal model did not converge; to simplify the model, the high correlation (1.0) between the random intercepts for item and the by-item random slopes for condition was removed. The difference in explained variance between the maximal model and the simplified final model was not significant.

Formula: Accuracy ~ TOWRE + Condition + (1 | Subject) + (0 + Condition | Item) + (1| Item) (Family = binomial)

	Estimate	***SE***	***z***	***p***
Intercept	-1.6	1.91	-0.84	.402
TOWRE	-0.02	0.02	0.98	.329
Condition	-2.74	0.51	-5.34	<.001

3.3.4 Masked priming lexical decision task

3.3.4.1 Main Model

Formula: (-1/RT) ~ TOWRE + Condition * Session * Relationship + (Session * Relationship | Subject) + (Condition * Session * Relationship | Item)

Fixed effects

	Estimate *(10^{-6})*	***SE*** *(10^{-6})*	***t***
Intercept	310.3	233.9	1.33
TOWRE	-13.4	2.015	-6.65
Condition	38.8	63.64	0.61
Session	-32.21	33.18	-0.97
Relationship	-29.53	32.24	-0.92
Relationship x Session	70.31	51.81	1.36
Relationship x Condition	-7.80	57.9	-0.16
Session x Condition	-149.9	66.93	-2.24
Relationship x Session x Condition	-210.4	100.4	-2.10

3.3.4.2 Pairwise comparisons to explore the three-way interaction

Four-exposures condition only:

Formula: (-1/ RT) ~ TOWRE + Session * Relationship + (Session* Relationship | Subject) + (Session * Relationship | Item)

	Estimate *(10^{-6})*	***SE*** *(10^{-6})*	***t***
Intercept	1347	556.1	2.42
TOWRE	-22.49	4.70	-4.53
Session	-101.6	37.68	-2.70
Relationship	-21.61	37.85	-0.57
Session x Relationship	-19.99	67.5	-0.30

Twelve-exposures condition only:

Formula: (-1/ RT) ~ TOWRE + Session * Relationship + (Session* Relationship | Subject) + (Session * Relationship | Item)

	Estimate *(10^{-6})*	***SE*** *(10^{-6})*	***t***
Intercept	100.5	221.7	0.45
TOWRE	-11.78	1.851	-6.36
Session	43.34	50.21	0.86
Relationship	-26.02	48.44	-0.54
Session x Relationship	176.8	70.15	2.52

Pre-exposure only for twelve-exposures condition:

Formula: (-1/RT) ~ TOWRE + Relationship + (Relationship | Subject) + (Relationship | item)

	Estimate *(10^{-6})*	***SE*** *(10^{-6})*	***t***
Intercept	97.95	225.5	0.43
TOWRE	-11.9	1.899	-6.27
Relationship	-96.36	53.54	-1.80

Post-exposure only for twelve-exposures condition:

Formula: (-1/RT) ~ TOWRE + Relationship + (Relationship | Subject) + (Relationship | item)

	Estimate *(10^{-6})*	***SE*** *(10^{-6})*	***t***
Intercept	323.7	429.5	0.75
TOWRE	-13.46	3.61	-3.73
Relationship	83.57	69.89	1.20

Appendix 4

Appendix 4.1

Word lists read by children in the non-semantic conditions

For each novel word, the list was laid out in three columns on an A5 page

List A

1. **allay**

asked, happy, going, flown, looked, now, fears, replied, flying, funny, plane, allay, chimed, help, told, scared, suggested, looking, fear, sounds, try, happier, allay, before, said, asked, aunt, said, explained, wrong, cousins, try, scared, visit, flying, allay, never, glad, told, admitted, tell, like, know, asked, journeys, insisted, parents, all, others, talked, fears, allay, scared, explained, fears, Sweden, today, flown, agreed, happy, fun, uncle

2. **bruit**

gossip, show, angrily, asked, told, fight, looked, stick, asked, worried, asked, same, going, said, bruit, saying, met, bruit, longer, heard, mean, wonder, school, bruit, started, talking, bad, gossip, today, true, explained, together, replied, bruit, said, look, gossip, nasty, rest, always, shouted, said, only, true, make, said, cares, firmly, today, people

3. **abort**

keep, tasty, one, bucket, plan, wave, headquarters, shower, said, headquarters, sent, wiping, house, green, hit, abort, tree, mission, ran, spy, surprise, today, mission, need, hid, moment, plan, objected, abort, called, pair, rest, tree, bucket, bushes, shouted, arms, face, shocked, plans, holding, agreed, house, watch, sight, snuck, back, house, window, looked, tree, sneak, jelly, abort, find, front, rest, mission, trying, garden, some, revenge, appeared, abort

4. **envoy**

ask, bad, houses, break, party, returned, arriving, news, envoy, organise, headquarters, cheered, lovely, bring, wanted, parents, all, accident, help, said, bike, tomorrow, poor, gone, party, envoy, thought, bones, help, go, forward, school, party, late, party, envoy, house, ask, need, really, luckily, preparing, today, house, wondered, surprise, envoy, friend, started, children, idea, looking, party, idea, friends, said, home, cheer, surprise, suggested, great, help, sent, wish, permission, come, quickly, stay

5. **felon**

criminals, life, yard, very, said, prison, America, others, card, explained, time, mean, works, impressed, felon, said, cousin, said, spiders, met, prison, spiders, all, bad, works, felon, scared, wish, shouted, cousin, America, asked, real, meet, today, said, admitted, felon, like, showed, lives, scared, interesting, really, think, met, visit, wonder, conversation, back, explained, felon, asked, interesting, grinned, said, today, think

6. abraid

idea, said, scare, night, grinned, camping, noises, morning, tents, hosepipe, tent, looked, night, freaky, boys, time, abraid, last, laughed, tell, fast, scare, time, making, revenge, sure, scared, abraid, first, boys, girls, trying, make, said, inside, said, asleep, girls, all, called, woke, night, faces, boys, separate, stir, even, abraid, morning, garden, realised, revenge, water, tried, go, laughing, spraying, wonder, fine, started, abraid, said, next, time, girls, boys

7. propel

rowboat, umbrella, propel, work, slowly, plank, rowing, look, narrow, holding, row, time, worked, try, propel, suggested, forward, tree, looked, tied, exclaimed, wooden, sight, old, forward, arm, look, boat, shore, propel, boat, climbed, plank, moving, paddles, go, decided, useful, asked, urged, discovered, using, replied, paddle, today, used, trip, very, lake, line, rowboat, use, going, think, started, long, lake, like, small, propel, boat, nearby, long

8. incite

here, whistle, incite, first, end, finished, wish, against, enough, chosen, playground
close, cheered, other, let, playground, battle, day, first, playground, glad, suggested
agree, blew, sunny, fastest, first, race, suggested, pick, battle, runner, race, incite, arrived, week, incite, managed, winners, playground, battle, trying, teams, battle, all, replied, said, shouted, race, battle, incite, race

9. pillage

used, like, order, battles, give, houses, Viking, today, project, used, pillage, boring, working, huge, try, pillage, Vikings, storm, tried, different, looking, Vikings, presentation, Vikings, classroom, things, famous, food, suppose, pretend, England, dress, boxes, books, presentation, library, school, acted, Club, suggested, Vikings, complained, special, finally, school, lunch, offered, money, take, class, asked, Vikings, continued, pillage, try, houses, kings, costumes, read, take, came, pillage, plan, success, act, suggested

10. secrete

dig, later, sure, box, spot, saw, buried, coin, secrete, suggested, tree, like, children, away, met, photo, back, away, liked, movie, decided, idea, figure, go, ten, pretty, treasure, initials, look, action, secrete, shoe, box, away, wondered, put, explained, secrete, hour, back, pact, all, box, stone, meet, box, brought, buried, treasures, today, treasures, come, forest, tree, edge, years, carved, special, old, secrete, forest, headquarters, made, box, home, mark, treasures, away

List B

1. croup

showed, fur, light, mum, clean, brush, calmly, started, favourite, mum, part, paying, passed, brushing, fur, croup, shiny, favourite, today, looked, soft, horse, particular, croup, very, owned, find, stood, friendly, looked, very, showed, mum, like, tail, replied, horses, end, called, croup, back, attention, horse, visiting, quite, croup, asked, go, teach, brush, said, brushed, content, brush, fur, showed, know, brown, horse, close, all, strokes, mum, brush

2. pivot

stop, found, mood, work, urged, painters, silly, dizzy, things, watching, pivot
assignment, started, assignment, chair, play, work, getting, called, finish, quickly, find, house, room, see, stop, started, French, yelled, school, French, said, never, making, said, tomorrow, just, game, spinning, pivot, make, like, meeting, complained
about, together, internet, dizzy, faster, painters, started, computer, dizzy, day, fast, chair, today, pivot, desk, work, look, pivot

3. antic

left, gloopy, performed, antic, headquarters, jelly, enemies, jelly, bottom, revenge, today, antic, wellies, boss, carried, time, food, antic, own, talking, house, asked, announced, came, stuck, prank, headquarters, loudly, suggested, tree, another, come, went, wait, think, pot, big, asked, look, snuck, hid, time, fill, played, cooked, wellies, weapon, green, wailed, jelly, one, goo, realised, house, sitting, gloopy, antic, bushes

4. wield

wield, fight, one, boss, bombs, picked, took, water, enemies, burst, wield, bomb, fight, playground, bomb, shouted, end, water, grass, gun, make, fight, found, asked, note, wet, very, saw, cheered, flying, threw, chosen, filled, three, look, all, together, showed, picked, meet, yelled, legs, went, fell, shooting, wield, door, wield, go, two, dodged, hit, water, balloons, water, bombs, three, wet

5. abode

even, abode, uncle, wore, entrance, replied, today, old, give, knight, walls, see, standing, members, wall, try, house, afraid, armour, uncle, end, big, hallway, family, visit, door, abode, showed, hope, disappointed, big, greeted, family, hall, large, enjoyed, abode, explained, great, stepped, furniture, tour, uncle, things, years, suit, mansion, said, offered, paintings, lived, abode, looked, big, mansion, welcome, asked, portraits, bit, interesting, visiting, old, uncle

6. suture

bad, suture, brave, put, scared, concerned, very, looked, cut, really, properly, skateboard, suture, gathered, started, replied, stitches, feel, take, suture, go, street, numb, happened, said, hurt, replied, doctors, asked, mum, hand, rock, realised, today, fell, made, around, playground, bandage, scared, cut, hand, asked, said, explained, asked, doctor, hospital, arrived, hand, tripped, suture, bandage, brave, heal, let, hand, doctor

7. warble

curly, song, interesting, singer, head, cheered, quivering, headquarters, aunt, Italian, house, aunt, voice, opera, sing, broke, tower, job, high-pitched, heads, warble, wore, language, cake, warble, aunt, makes, trying, asked, invited, suggested, sounded, good, size, try, hair, funny, huge, voice, attempt, asked, tea, famous, pouring, aunt, looking, voice, explained, aunt, silly, warble, tell, sang, copy, foreign, pitch-black, shook, piled, Sunday, warble, laughed, song, person, twice, cake, good, tea

8. palate

prank, said, really, tried, make, look, cake, food, talking, replied, rarely, snails, cream, made, cake, spicy, eaten, sat, sauce, suit, special, magnificent, palate, food, France, food, use, asked, cake, plan, special, prank, spicy, sure, surprise, gives, idea, like, continued, really, idea, exclaimed, worst, mayonnaise, funny, yummy, disgusting make, added, ever, palate, grinned, chilli, cake, today, hide, chimed, suits, barbecue, kitchen, sardines, sauce, holiday, food, palate, chocolate, palate, suit

9. dudgeon

write, bet, thing, said, high, laughed, dudgeon, ugly, goo, nasty, message, meet, write, calm, revenge, all, shouted, message, today, asked, revenge, better, girly, wrote, slow, thing, members, gloopy, dudgeon, high, suggested, chalk, large, house, dudgeon, pavement, written, said, butterflies, called, draw, transform, said, exclaimed, still, think, grumpily, letters, pavement, high, sentence, try, said, suggested, nasty, dudgeon, all

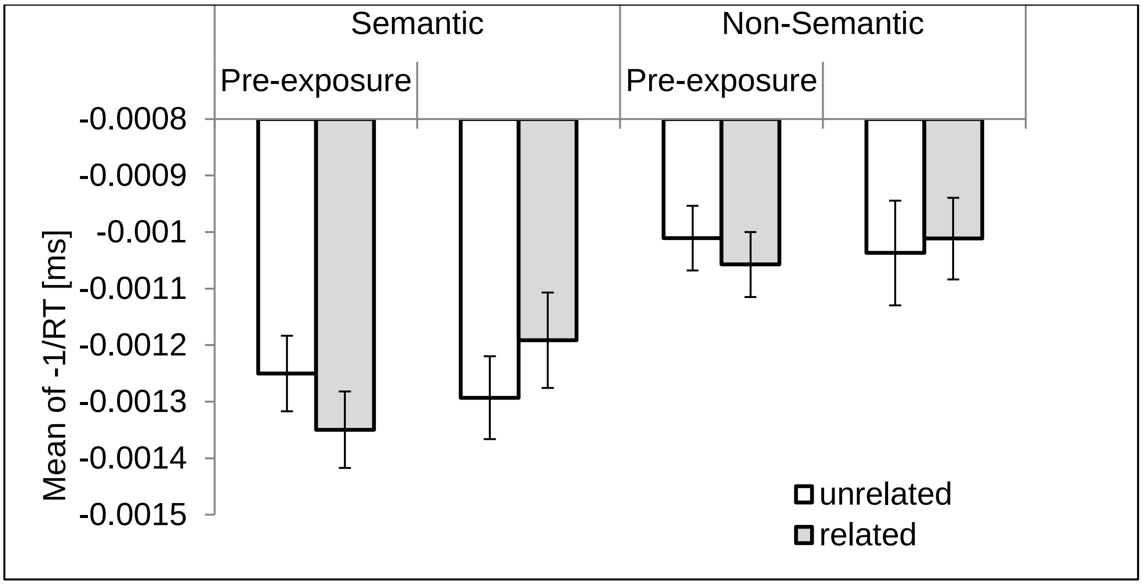
10. undermine

prisoners, own, far, tale, today, made, tunnel, tried, exciting, tell, asked, no, undermine, reason, email, raining, caught, manage, surprised, cousin, undermine, said, managed, explained, said, edge, floor, hear, end, very, today, one, good, house, meeting, undermine, tried, cell, America, prison, guards, way, told, tools, all, build, prisoners, curious, escape, said, concrete, prison, wondered, floor, works, prison, escape, story, guard, prison, undermine

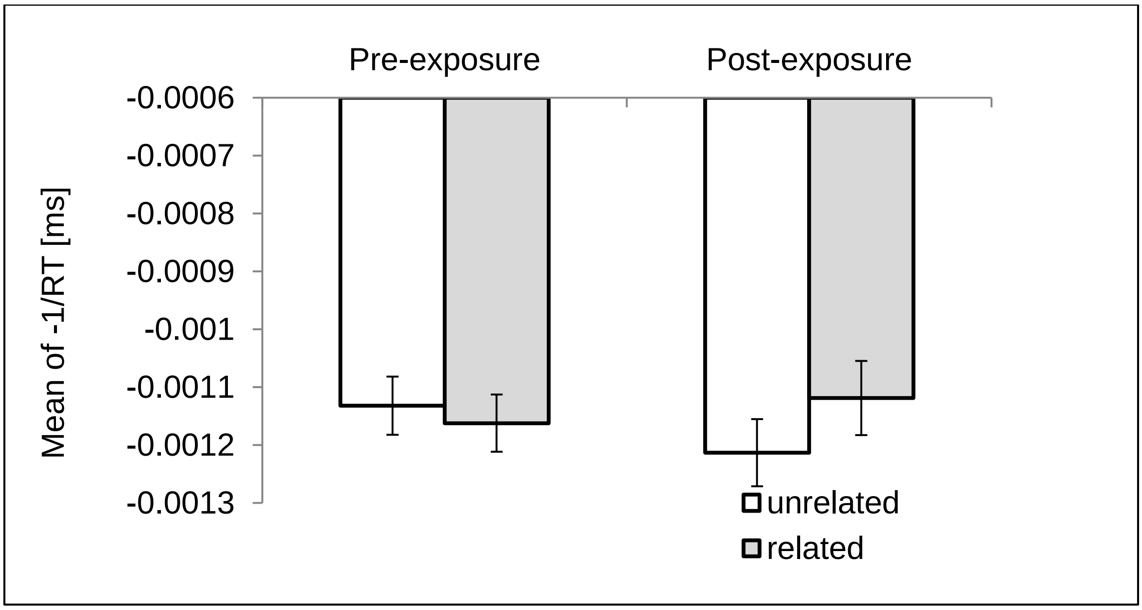
Appendix 4.2

Inverse transformed data for masked priming lexical decision (Experiment 3.1)

4.2.1 Data split by exposure condition. Error bars show 95% confidence intervals.



4.2.2 Data collapsed across exposure conditions. Error bars show 95% confidence intervals.



Appendix 4.3

Notes. 'Condition' refers to semantic vs. non-semantic.

4.3.1 Reading accuracy during exposure phase

Formula: Accuracy ~ TOWRE + Condition + (1 | Subject) + (Condition | Item) (Family = binomial)

Fixed effects:

	Estimate	***SE***	***z***	***p***
Intercept	-0.03	1.27	0.08	.978
TOWRE	0.02	0.01	0.62	.078
Condition	-0.64	0.35	-1.85	.064

4.3.2 Definition knowledge

Formula: Accuracy ~ TOWRE + Condition + (1 | Subject) + (Condition | Item) (Family = binomial)

Fixed effects:

	Estimate	***SE***	***z***	***p***
Intercept	-1.73	1.43	-1.21	.230
TOWRE	0.01	0.01	0.62	.538
Condition	-4.62	0.83	-5.541	<.001

4.3.3 Orthographic decision

Formula: Accuracy ~ TOWRE + Condition + (1 | Subject) + (Condition | Item) (Family = binomial)

Fixed effects:

	Estimate	***SE***	***z***	***p***
Intercept	-5.38	2.70	-1.99	.046
TOWRE	0.07	0.03	2.69	.007
Condition	-0.72	0.88	-0.82	.413

4.3.4 Spelling

Formula: Accuracy ~ TOWRE + Condition + (1 | Subject) + (Condition | Item) (Family = binomial)

Fixed effects:

	Estimate	***SE***	***z***	***p***
Intercept	0.98	1.41	0.70	.486
TOWRE	0.01	0.01	0.91	.362
Condition	-0.43	0.70	-0.61	.545

4.3.5 Masked priming lexical decision

4.3.5.1 Main model

Formula: (-1/RT) ~ TOWRE + Condition * Session * Relationship + (Session * Relationship | Subject) + (Condition * Session * Relationship | Item)

Fixed effects:

	Estimate *(10^{-6})*	***SE*** *(10^{-6})*	***t***
Intercept	179.40	212.50	0.84
TOWRE	-12.17	1.89	-6.43
Condition	55.68	82.10	0.68
Session	20.77	22.09	0.94
Relationship	-37.33	44.14	-0.04
Relationship x Session	130.90	45.02	2.91
Relationship x Condition	20.85	62.55	0.33
Session x Condition	-37.33	44.14	-0.85
Relationship x Session x Condition	-96.14	131.90	-0.73

4.3.5.2 Pairwise comparisons to explore prime-target relationship x test session interaction

Pre-exposure only:

Formula: (-1/RT) ~ TOWRE + Relationship + (Relationship | Subject) + (Relationship | item)

Fixed effects:

	Estimate *(10^{-6})*	***SE*** *(10^{-6})*	***t***
Intercept	195.5	215	0.91
TOWRE	-12.41	1.93	-6.44
Relationship	-64.24	33.35	-1.93

Post-exposure only:

Formula: (-1/RT) ~ TOWRE + Relationship + (Relationship | Subject) + (Relationship | item)

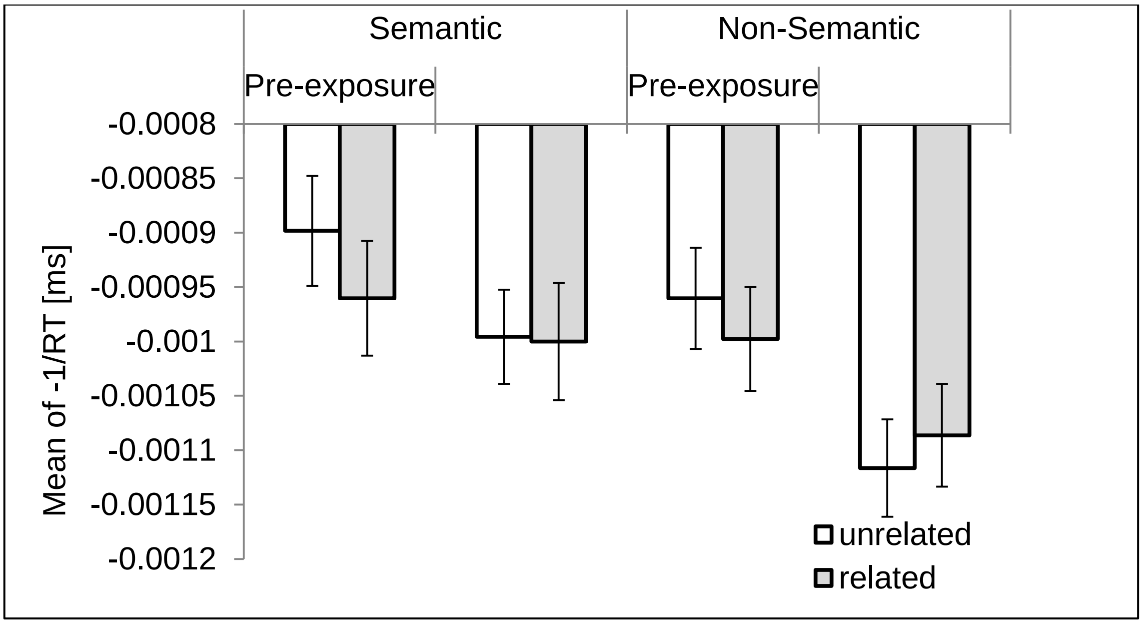
Fixed effects:

	Estimate *(10^{-6})*	***SE*** *(10^{-6})*	***t***
Intercept	36.37	271.1	0.13
TOWRE	-10.76	2.436	-4.42
Relationship	59.73	43.67	1.37

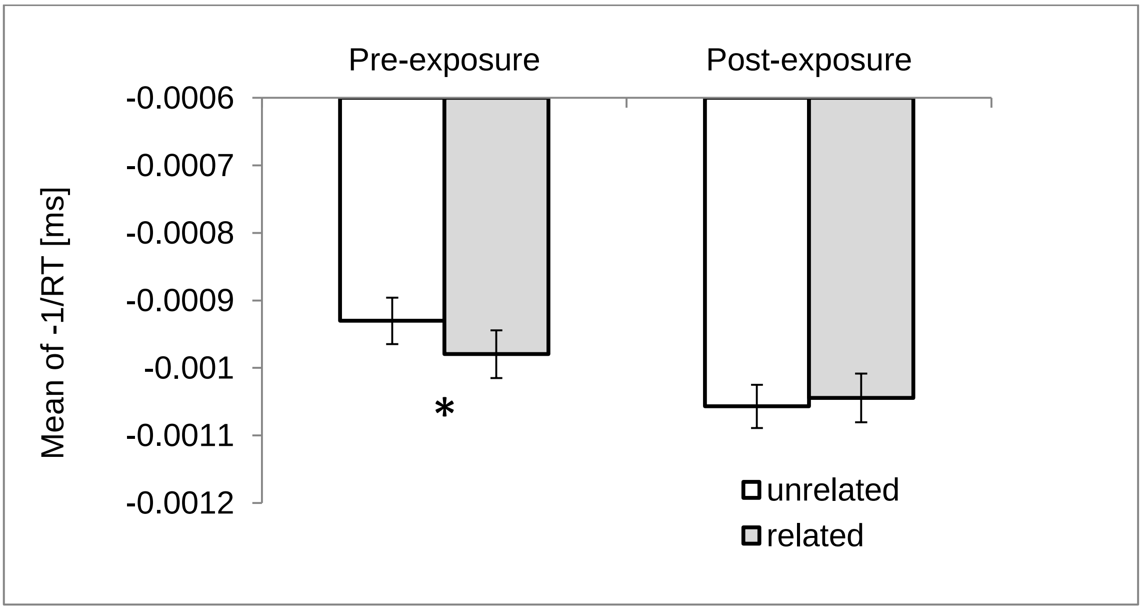
Appendix 4.4

Inverse transformed data for masked priming lexical decision for (Experiment 3.2)

4.4.1 Data split by exposure condition. Error bars show 95% confidence intervals.



4.4.2 Data collapsed across exposure conditions. Error bars show 95% confidence intervals.



Appendix 4.5

4.5.1 Definition knowledge

Formula: Accuracy ~ TOWRE + Condition + (1 | Subject) + (Condition | Item) (Family = binomial)

Fixed effects:

	Estimate	***SE***	***z***	***p***
Intercept	-6.34	2.38	-2.67	.008
TOWRE	0.05	0.02	2.47	.013
Condition	-2.85	0.69	-4.16	<.001

4.5.2 Reading accuracy at post-test

Notes. The maximal model did not converge; to simplify the model, the high correlation (1.0) between the random intercepts for item and the by-item random slopes for condition was removed. The difference in explained variance between the maximal model and the simplified final model was not significant.

Formula: Accuracy ~ TOWRE + Condition + (1 | Subject) + (0 + Condition | Item) + (1| Item) (Family = binomial)

Fixed effects:

	Estimate	***SE***	***z***	***p***
Intercept	-1.58	1.74	-0.91	.363
TOWRE	0.03	0.02	2.25	.024
Relationship	-0.03	0.47	-0.06	.951

4.5.3 Spelling

Formula: Accuracy ~ TOWRE + Condition + (1 | Subject) + (Condition | Item) (Family = binomial)

Fixed effects:

	Estimate	***SE***	***z***	***p***
Intercept	-5.81	1.39	-4.19	<.001
TOWRE	0.06	0.01	4.95	<.001
Relationship	0.29	0.39	0.73	.465

4.5.4 Masked priming lexical decision

4.5.4.1 Main model

Formula: (-1/RT) ~ TOWRE + Condition * Session * Relationship + (Session * Relationship | Subject) + (Condition * Session * Relationship | Item)

Fixed effects:

	Estimate *(10^{-6})*	***SE*** *(10^{-6})*	***t***
Intercept	-763.40	230.70	-3.31
TOWRE	-1.97	1.90	-1.04
Condition	-89.23	65.61	-1.36
Session	-99.28	23.53	-4.22
Relationship	-19.17	16.73	-1.15
Relationship x Session	62.77	28.70	2.19
Relationship x Condition	18.49	33.42	0.55
Session x Condition	-38.23	51.58	-0.74
Relationship x Session x Condition	20.47	62.62	0.33

4.5.4.2 Pairwise comparisons to explore prime-target relationship x test session interaction

Pre-exposure only:

Formula: (-1/RT) ~ TOWRE + Relationship + (Relationship | Subject) + (Relationship | item)

	Estimate *(10^{-6})*	***SE*** *(10^{-6})*	***t***
Intercept	-440.60	285.80	-1.54
TOWRE	-4.26	2.36	-1.80
Relationship	-49.15	20.08	-2.45

Post-exposure only:

Formula: (-1/RT) ~ TOWRE + Relationship + (Relationship | Subject) + (Relationship | item)

Fixed effects:

	Estimate *(10^{-6})*	***SE*** *(10^{-6})*	***t***
Intercept	-837.40	241.00	-3.48
TOWRE	-1.79	1.99	-0.90
Relationship	13.41	20.37	0.66

Appendix 5

Appendix 5.1

Additional stories read in sessions 2 and 3 in the diversity condition

Session 2:

List A

1. A message on the pavement

Today Jimmy called all the members of the Caterpillar Club to meet at his house. He was not happy. Someone had written a sentence with chalk in large letters on the pavement. It said: 'Caterpillars are slow and ugly!' "I bet the Gloopy Goo Gang wrote that!" Jimmy exclaimed. "What a nasty thing to do," Rachel said. "What shall we do?" "I'm furious!" Billy shouted. "And only revenge will allay my anger!" "If revenge is the only way to allay your fury, write them a message too!" Sara suggested. "What would we write?" Jimmy asked. "Gloopy goo is ... gloopy?" Everyone laughed, except Billy, who was still fuming. "Let's draw butterflies all over their pavement, to show them that caterpillars can transform into anything!" Rachel suggested. "A girly thing like that won't allay my anger," Billy said grumpily. "While you try to think of something that will allay it, I will go with Rachel" Sara said.

2. Aliens?

Today the Caterpillar Club was meeting at their headquarters. Sara was very excited when she arrived. "Guess what," Sara said. "Today at school, I overheard Mary bruit about an interesting rumour. She told everyone that a spaceship landed in the field behind her house.” “Let's go check out if Mary really had something to bruit about!" Billy called. The Caterpillar Club went to investigate the rumour about the spaceship that Sara heard Mary bruit about. There they found a giant beanshaped object in the field. "That's not a spaceship," Sara said. "That's a hot air balloon. It's just shaped like a spaceship. I wonder what they are doing with it." "Look!" exclaimed Rachel. "There is a sign saying they are shooting a movie." "Maybe we can be in it!” Billy cheered. "First, let's tell Mary what her spaceship is really about, so she won't bruit about it anymore,” Jimmy said.

3. Who’s the boss

Even the best club in the world sometimes argues. “Why is Jimmy the boss?” asked Billy. “Because I had the idea for the Caterpillar Club,” said Jimmy. “So?” said Billy. “Obama is the president of America. But America wasn’t his idea!” Rachel and Sara laughed. “Who else should be the boss? Rachel and Sara can’t be because they are girls,” Jimmy said. “Now that is not true!” Sara interjected. “Girls can be bosses.” “Would you three please abort your argument?” Rachel pleaded. “Yes, let's abort this silly discussion," Sara agreed. "Maybe we just shouldn’t have a boss.” “We can't abort this discussion. Someone has to be in charge," Jimmy replied. "Then we should take it in turns," Sara suggested. "You can be the boss for another week, and then we will swap." Eventually Billy and Jimmy agreed to abort their argument.

4. A present for Rachel's Dad

Today the Caterpillar Club was helping Rachel make a present for her dad. "My dad is going to France tomorrow for two weeks, so he won't be here for his birthday. He works as an envoy for the British government." "What is he going to do in France?" Jimmy asked. "As an envoy, he has to deliver top secret messages and talk to important people," Rachel explained. "Being an envoy sounds like an exciting job!" Billy said. So they all got going helping Rachel with her present. She was making a photo album. She had collected lots of pictures of her family. The Caterpillar Club was cutting them out in pretty shapes and gluing them into the album. "I am very proud of my dad being an envoy," Rachel said. "But I wish he didn't have to be in a foreign country on his birthday!" "This album will make him think of you. It's a great present!" Sara said.

5. Capture the flag

Today the Caterpillar Club were playing Capture the Flag against the Gloopy Goo Gang. The playground was divided into two sides. You could only be tagged on the other team's side. Then you had to go to that team's prison. Each team hid a flag somewhere on their side. You had to find and capture the other team's flag. Ollie from the Gloopy Goos dared to run into the Caterpillars' side, but Billy managed to tag him. "You need to go to our prison!" Billy shouted. "You are a felon!" Rachel was guarding the Caterpillars' prison under the slide. "I will make sure this felon doesn't escape," she said. Sara got tagged by the Gloopy Goo Gang and was sitting in their prison as a felon. Luckily, Jimmy managed to free her and she was no longer a felon. In the end, the Caterpillar Club managed to capture the Gloopy Goos flag from the top of the climbing frame.

6. A thundery surprise

The Caterpillar Club was spending a lazy day in Jimmy's garden. Sara and Jimmy were playing Frisbee. Rachel was drawing a flower in her sketch book. Billy was having a nap in a hammock hanging from a tree. They didn't notice that dark clouds were appearing in the sky. Suddenly, a loud thunder clap surprised them. The noise was loud enough for Billy to abraid. When he opened his eyes and started to abraid properly, he was so startled he fell out of the hammock. He got back on his feet and rubbed his eyes. "Thunder is a good alarm clock for Billy," Jimmy said laughing. "I was in the middle of a dream when the thunder made me abraid!" Billy said. "Well I am glad it made you abraid," Rachel said, "Help me carry my things inside, it's starting to rain!" More thunder, lightning and rain followed. The Caterpillar Club ran to seek shelter inside.

7. What a ride!

Today the Caterpillar Club was in town. They were watching a clown performing a juggling act. He was standing on a strange vehicle. “It’s a bathtub on wheels!” Sara exclaimed. After a while the clown sat down to take a break. “Excuse me,” said Jimmy. “Does your tub actually drive?” “Oh yes,” replied the clown. “It isn’t very fast, but it gets me where I need to go.”
“How does it work?” asked Sara. “I put an engine in to propel it forward,” the clown explained. “A small one is enough to propel the tub forward.” “Please can you show us?” Rachel begged excitedly. “Sure,” the clown replied, “and I can take one of you along for the ride! Any more and the engine won't be able to propel us.” Of course they all wanted a go, so one after the other they rode in the bath tub around the market square. "I can't believe how fast this engine can propel us forward!" Billy yelled.

8. The diary

Today Billy was very excited when he arrived at the Caterpillar headquarters. “Look what I found on the floor in the school corridor,” he said holding up a notebook. “It’s Karina’s diary!” Karina was a member of the Gloopy Goo Gang. “Did you read it?” Rachel asked. “Not yet,” Billy said. “But I think we should try to find out what they are up to!” “I think you are trying to incite us to do something bad,” Rachel said. “The Gloopy Goos are our enemies, but we can’t be horrible to them,” Sara agreed. “Why do you think I am trying to incite you to do something horrible?” Billy asked. “You are trying to incite us to break Karina’s privacy. A diary is personal,” Rachel said. “I would be very upset if someone read my diary,” Sara added. “Fine,” Billy agreed. “I won’t try to incite you to read it any longer. I will put it in her locker.”

9. What is your strength?

Today the Caterpillar Club were sitting on the grass outside their headquarters. Everyone was bored. “We should teach the Gloopy Goo gang a lesson,” Jimmy proposed. “What kind of lesson?” asked Rachel. “We should show them our strengths,” Jimmy explained. “Let’s pillage the Gloopy Goo headquarters!” said Billy. “What do you want to steal from them?” Sara asked. “I’ll decide what to pillage when I’m there!” Billy replied. “I don’t think trying to pillage is our strength,” Rachel objected. “That’s right,” Jimmy agreed. “Let’s think about what each of us is good at before we decide to pillage them.” “Sara is very clever,” Rachel said. “Jimmy is fast and strong,” Sara went on. “I am fearless!” Billy declared. “And Rachel looks after everyone,” Jimmy finished. “We should work out a plan that uses all of our strengths,” Sara suggested. This kept the Caterpillar Club busy for the rest of the day.

10. Party intruders

Today Rachel was hosting a summer garden party. Her parents were helping her have a big barbecue. To make the party more special, Rachel had written on the invitations that everyone had to wear masks and fancy dress. She had invited lots of her friends. The party was going very well, when suddenly four children arrived in masks. Rachel wasn't sure who they were because the masks did well to secrete their eyes. It was two boys and two girls. "It's the Gloopy Goo Gang! They are trying to sneak into our party," Billy shouted. Billy was right. The masks had not been enough to secrete the faces. The Gloopy Goos took their masks off and laughed. "Didn't we do well to secrete our identity like this!" Harry yelled. "Maybe you tried to secrete who you are, but now we know. Get out!" Jimmy yelled back. The Gloopy Goo Gang laughed and ran away.

List B

1. Tricky tricks on Copper

Today the Caterpillar Club was getting a trick riding lesson on Copper, Rachel's mum's horse. "Trick riding is when you perform tricks and acrobatics on the back of the horse," Rachel's mum explained. "For example, two people can stand on the back of a horse near the croup at the same time. The person at the front stands where the saddle would be, and the person at the back stands on the horse's croup." "I want to try standing on Copper's croup!" Billy shouted. "We are going to start with something easy," Rachel's mum climbed on Copper. "Sit on the horse, then swing your legs to one side. Next you swing around to sit backwards to face Copper's croup." Copper started trotting. "Ho! Slow down Copper!" Rachel's mum called. "How is riding a horse backwards?" Rachel asked laughing. "Not as easy as forwards," Rachel's mum replied. The Caterpillar Club all had a go at performing the trick.

2. Playground attack

Today the Caterpillar Club was standing in the playground. They were discussing their plans for the afternoon. They did not know that the Gloopy Goo Gang was sneaking up on them from behind. Harry the boss started whistling to attract their attention. Jimmy was the first to pivot around. He saw something flying towards him and dodged. He yelled at the others to pivot around as well. The Gloopy Goo Gang was throwing balloons at them. The Caterpillar Club decided to pivot to the right and run behind the slide for cover. “Are these water bombs?” Rachel asked. She tried to pivot around to look at one behind her. One of the balloons hit the slide and exploded. It was filled with green gloopy goo. “It’s jelly!” Billy shouted. Finally the Gloopy Goo Gang ran away. The Caterpillar Club came out from behind the slide. There was green jelly everywhere. “We need to get them back for this!” Jimmy shouted. The others agreed.

3. Fun with Max

Today the Caterpillar Club was in Billy's garden, playing with his dog Max. Max was running in and out of the house. He brought out his ball and put it down in front of Sara. Sara picked it up and threw it for Max to catch. Max carried the ball back into the house. “What silly antic are you getting up to?” Rachel called as Max came outside again. Max had brought out a shoe. "That is one of Mum's," Billy said. "You better put it back or she will tell you off for your naughty antic." Billy took the shoe off Max, but it didn't take long for Max to come up with a new antic. He brought out a cushion from the sofa. "He wants to lie down on it on the terrace," Billy explained. "But I will get told off if I let him get away with this antic!" In the end, Max went back to playing with his ball.

4. The green door

Today the Caterpillar Club were painting the garage door at Sara’s house. Sara’s dad had offered them money for ice cream in return. Soon they were busy painting the door green. “You missed a spot there. I knew girls aren't good at this kind of stuff!” Billy said to Sara. He used his brush to wield it at the spot. “Take that back or I’ll paint you in a minute!” Sara threatened. She started to wield her brush at Billy. “Don’t wield that brush at me!” Billy yelled. “Stop using the brushes to wield them at each other,” Rachel interfered. When Sara’s dad saw the finished garage door, he didn’t seem happy. “Oh dear,” he sighed, “I gave you the wrong paint! Your mother wants a blue garage door!” Everyone moaned. “Not to worry,” the dad said, “Tomorrow you can paint blue over the top. You can get some extra ice cream.” That was a great idea.

5. Help for Mary

Today the Caterpillar Club was helping their friend Mary. Mary was moving house and they were helping her pack up the things in her room. "I am glad I am only moving to the other side of town." "Why are you moving at all?" Billy asked. "I am going to get a little sister. We need a bigger abode so that there is space for her," Mary explained. "Have you been to your new abode?" Rachel asked. "Yes, my parents showed it to me. It is really nice. It will be the first time I'm living in an abode with a garden." The Caterpillar Club helped Mary carefully wrap up all her possessions and put them into boxes. "Changing abode is a lot of work." Jimmy said. "You have so many things!" "Thank you so much for helping me," Mary replied. "This would take forever if I did it by myself." By the end of the day, they had packed everything.

6. Jimmy's new bandage

Today Jimmy arrived at the Caterpillar headquarters with a fresh big plaster on his arm. A few weeks before he had fallen off his skateboard and cut his arm. “How is your arm?” Rachel asked. "Did they have to suture the cut again? "No," Jimmy replied. "I just had to go to the doctor to have the stitches removed. The cut has healed now." “Is there going to be a scar from where the doctor had to suture the cut?” Billy asked. "I think so," Jimmy replied. "I don't think it's possible to suture a cut without leaving a scar. But I don't mind that. I think it will look cool." "I agree!" Billy exclaimed. "Well it that case I am glad it was you they had to suture. I wouldn't like a scar so much," Sara said. Jimmy shrugged. The Caterpillar Club went on to play a game of tag.

7. Treasure in the woods

Today the Caterpillar Club decided to go on a treasure hunt in the woods. However, it was difficult to decide where to start looking. "This was a terrible idea," Billy complained. "Treasure is hidden by pirates or smugglers who travel the sea. We are never going to find any!" "Listen to that!" Sara exclaimed. They heard a beautiful sound. It was a bird singing somewhere very close to them. Sara started to walk toward the bush where she first heard the bird warble so beautifully. She carefully peeked between the twigs where she heard it warble. There was a bird's nest with at least five or six baby birds squeaking with open beaks. The others gathered around Sara to listen to the father of the babies warble his beautiful song. The Caterpillar Club listened and watched quietly. They did not want to scare the birds. "I told you we would find some treasure. I could listen to him warble all day," Sara whispered.

8. An interesting dish

Today the Caterpillar Club was at Rachel's house for dinner. Rachel's mum was trying out a new recipe. It was a German stew. "I don't think I have ever eaten anything German," Jimmy said as they sat down. "I don't have a very experienced palate." "Me neither," said Billy. "I just like sausages, chips and beans!" "I like to try out new things," Sara said. "You never know, the food might turn out to really suit your palate." Rachel's mum served the stew. It was made with chicken, mushrooms, and small, round, green vegetables. "What are these?" Sara asked. "They look like strange peas." "These are capers," Rachel's mum explained. The Caterpillar Club all tried the stew. It was a new experience for everyone's palate. "I am not sure about the capers," Billy said. "I really like them!" Rachel disagreed. "Well, wouldn't it be boring if we all had the same palate?" Sara wondered. The others nodded.

9. Sara's brother

Today Sara arrived at the Caterpillar Club headquarters looking serious. "I'm in quite some dudgeon about my big brother," Sara fumed. "What has he done?" Billy asked. Sara sat down next to the others. "Well, we were planning to go on a family trip to the zoo this weekend," Sara said. "But now he has convinced my dad to go karting instead." "I don't know why you are in dudgeon about this," Jimmy joined in. "I would much rather go gokarting than go to the zoo." "Me too!" Billy agreed. "You two are just like my brother!" Sara said in high dudgeon. "I have been looking forward to the zoo for weeks!" "Does your brother know that you are in a dudgeon about this?" Rachel asked. "Not really," said Sara. "He will say I am just scared I will lose a gokart race against him." "I think you should try to talk to him," Rachel suggested. "Maybe I will," Sara agreed.

10. Hidden passages?

Today the Caterpillar Club were working on a school project about the London Underground. “Isn't it amazing that they managed to build all these tunnels for the trains?" exclaimed Sara. "They must have had to undermine so many streets and buildings." "I am impressed they were able to do that 150 years ago," Jimmy added. "It must have been hard work to undermine all of London without any big machines." "I don't think I could undermine a street with just a shovel," Billy agreed. "I wish we could build our own hidden underground passage," Sara wondered. "It could lead straight into our headquarters." "We would have to undermine our garden for that," Jimmy said and laughed. "I don't think my parents would let us do that!" "Well, the point of a hidden passage is that no one else knows about it!" said Sara. The Caterpillar Club continued to imagine their own network of secret tunnels.

Session 3

List A

1. A parcel from Grandma

Today the Caterpillar Cub met at Rachel's house. Her grandma had sent a parcel that was addressed to all of the Caterpillar Club. It was a big, heavy box. Finally, everyone had arrived and they sat down around the parcel. "I really want to have a look inside to allay my curiosity," Rachel said excitedly. She opened the parcel at the top. There were several things wrapped in newspaper, as well as a note. It said: 'I heard about your club. Here are some things from my cellar that might be useful. Grandma.' "Let's unwrap everything!" Billy said. "I can't wait to allay my curiosity." "Me too!" Sara said. "If I don't allay my curiosity soon, I am going to burst!" So the Caterpillar Club started unwrapping all the things at once to allay everyone's curiosity. They found the following things: A torch, a compass, binoculars, a guide to edible berries and mushrooms, and a pocket knife.

2. Harry's hedgehog

Today everyone had breakfast at Sara's house. Sara's dad was reading the newspaper. "Listen to this," he said. "Your friend Harry is in the newspaper! " "He is the boss of the Gloopy Goo Gang," Sara replied. "He is not our friend" "Either way," her dad continued. "It says that he rescued a hedgehog. I am not surprised that the newspaper chose to bruit about this. They love to write about local children and bruit about their achievements.” “We should try and rescue an animal, too,” Sara suggested. "Then they will bruit about us!" “Let’s find something bigger than a hedgehog for them to bruit about!” Billy yelled. “How about that dog over there?” Rachel pointed at a dog across a street. “That is Mrs Beaver’s dog,” Sara objected. “Look, she is calling him back. We can’t just rescue him from her.” The others laughed. They decided to keep their eyes open for any animals in trouble.

3. Frightful flight

Today the Caterpillar Club were out in a field playing with a model aeroplane that Billy received for his birthday. It came with a remote control that let you steer its flight. Billy flicked the switch on the remote control and the aeroplane took off into the air. Suddenly, a strong gust of wind swept over the field. The aeroplane drifted off its path. “It’s going to hit the tree!” Rachel cried. “Turn left! Maybe you should abort the flight.” Billy tried to abort the flight, but the wind was too strong. The aeroplane hit the tree top and disappeared between the leaves. “Now we really have to abort the flight!” Sara shouted. “Bring back the plane!” But the aeroplane was stuck. Luckily Rachel was excellent at climbing trees. She reached the aeroplane and freed it from the branches. “Let’s go somewhere without trees so we don't have to abort another flight,” she said when she was back on the ground. So the Caterpillar Club set off to search for an open space.

4. A boy visits

Today the Caterpillar Club were at the playground when a small boy ran towards them. "Are you the Caterpillar Club?" the boy asked. "I was sent as an envoy from the Gloopy Goo Gang." "Why are you working as an envoy for them?" asked Jimmy. "I want to join their gang. They told me I have to do a few things for them first," the boy explained. "They sent me as an envoy to bring you this message: The Gloopy Goo Gang challenges you to a game of chess. Do you accept?" "Sara is very good at chess," Jimmy said. "We accept." "Why do you want to join the Gloopy Goo Gang?" Rachel asked. "I want to be a member of a gang," the boy said. "Maybe you should start your own gang. Then you don't have to be an envoy," Sara suggested. "I'll think about it," the boy replied. "I've got to get back to the Gloopy Goo Gang now."

5. Where is Max?

Today Billy called an emergency meeting at his house. He was rather upset. "My dog ran away!" he exclaimed. "I must have left the back door open." "That was pretty silly," Jimmy said. "Don't look at me like I'm some kind of felon!" Billy said unhappily. "I didn't do it on purpose!" "Let's not waste time talking about who is a felon," Rachel said. "We need to go look for Max." The Caterpillar Club started looking in the streets. They shouted his name and asked people walking by, but no one had seen Max. "If we don't find him, I really will feel like a felon," Billy said miserably. "You are not a felon, and we will find him!" Rachel comforted him. Suddenly, Max shot around the corner and ran straight towards Billy, happy to see his owner. "There you are!" Billy cheered. "What adventures have you been up to?"

6. A dog's dream

Today the Caterpillar Club was at Billy’s house. It was a sunny day and they were playing with Billy’s dog Max. They were throwing a tennis ball around that Max kept chasing. After 20 minutes, everyone needed a break. Max trotted to a sunny spot on the terrace. He lay down and quickly fell fast asleep. The Caterpillar Club was watching him. He started wagging his tail and moving his legs in his sleep. “He is probably dreaming about chasing the ball,” Billy said. “Shh,” Rachel said. “I don’t want him to abraid! This is too much fun to watch.” “Don’t worry. Once he is asleep, he does not abraid for a long time, even if you want him to abraid,” Billy said. At that moment, Billy’s mum came outside with a tray. “I have some sausage sandwiches for you,” she called. That was enough for Max to abraid straight away. He jumped to his feet. The Caterpillars laughed.

7. Challenges

Today the Caterpillar Club received a note from the Gloopy Goo Gang. “Meet at our headquarters at 4pm today.” The Gloopy Goo gang was standing under their tree house with their arms folded and stern looks on their faces. “This might be a trap,” said Sara. “We should be careful." Billy pushed Jimmy to propel him forward to their enemies. "Why are you trying to propel me towards them?" Jimmy asked. “He is trying to propel you to talk to their boss,” Sara said. "No need to propel me like that. I'll do it," Jimmy said. “I suggest that we hold a competition to find out which is the better gang,” Harry the boss shouted. “What kind of competition?” Sara asked. “I propose that we each come up with some challenges to test the other gang,” Harry explained. “When should we have the competition?” Rachel asked. “Next weekend!” Jimmy suggested. “That will give us all time to prepare the challenges.” everyone agreed to this plan.

8. Sara in demand

Today the Caterpillar Club was meeting at a field next to a river. Sara arrived late. Everyone else was already cooling off their feet in the water. “Guess what!” Sara said. “Harry from the Gloopy Goo gang just spoke to me. He tried to convince me to leave the Caterpillar Club and join the Gloopy Goo Gang!” “I can’t believe Harry did that!” Rachel said outraged. “Why would anyone ever want to join the Gloopy Goo Gang?” Jimmy laughed. “Was he offering you anything to incite you to join them?” Billy asked. “Yes, he offered me a place as second in command to incite me,” Sara explained. “I don’t think anyone would be tempted by that. That would never incite me,” Rachel said. “You don’t want to join them, do you?” “Of course he did not manage to incite me. I told him I would never leave the Caterpillar Club,” Sara said. “And I certainly would never ever join the Gloopy Goo Gang!”

9. Rachel the chef

It was lunch time when the Caterpillar club returned from their adventures. “I’m hungry,” Billy moaned. “Let’s pillage our kitchen and see what we can find,” said Sara. They snuck into the house. "Let's pillage the fridge!" Billy suggested. They managed to pillage some bread and cheese for cheese on toast. Rachel buttered the bread and grated lots of cheese on every slice. “We need to put these in the oven now,” she explained. Rachel's mum came to help them with the oven. “This is taking ages,” Billy complained. “Let’s play some cards.” They started a game and forgot all about the cheese. “Something smells burnt,” Jimmy wondered after he won the first round. “The cheese!” Rachel called. There were four black slices in the oven that had been burnt. Luckily, there still was some bread and cheese left to pillage.

10. A suspicious object

Today Jimmy was very excited when he met the other Caterpillar members. "The Gloopy Goo Gang is building something in Harry's yard," he explained. "What are they building?" Rachel asked. "I am not sure. It looked a bit like a cannon. Maybe it's a weapon. We have to go and spy on them!" Jimmy replied. The Caterpillar Club went to Harry's house. "We need somewhere to secrete ourselves, so they can't see us," Sara said. "How about we secrete ourselves behind that row of bushes?" Rachel suggested. Everyone agreed that that was a good place to secrete themselves. But they didn't have to secrete themselves for long. "I know what they are building," Sara said, "That's not a cannon, it's a giant telescope! You need it to look at stars and planets that are very far away." “Better to be safe than sorry,” Jimmy said as they walked back to their headquarters.

List B

1. Beautiful Esmeralda

Today the Caterpillar Club was at a horse show with Rachel's mum. There were lots of different types of horses in all shapes and colours. They particularly liked a small, white, shaggy pony. His name was Joker and the owner let them stroke him. Rachel's mum took them to see some very expensive horses. They were tall and slender with very shiny fur. Rachel's mum particularly liked a brown one called Esmeralda. "Look at her beautiful thin neck!" Rachel's mum exclaimed. "And the shape of her croup is perfect. A horse shouldn't have a flat back. Ideally you want a nicely curved croup." "I see what you mean about Esmeralda's croup," Sara agreed. "None of the other horses has a croup shaped quite as nicely." Next, the Caterpillar Club went on to watch a horse riding performance. "Maybe I should learn to do that," Jimmy said. "The riders all look very strong and powerful!"

2. Run Rupert, run

Today the Caterpillar Club was at Billy's house. They wanted to see Billy's new hamster. "What's his name?" Rachel asked. "Rupert," Billy replied. Rupert was living in a lovely cage with a wheel to run in. "Look at him running in his wheel," Sara exclaimed, "Does he not get bored? All the wheel does is pivot around on the spot!" "I think he likes making the wheel pivot fast," Billy replied. "I wonder how he keeps up with it once it starts to pivot really fast." Billy took Rupert out of his cage. He caught him in his wheel, which continued to pivot around. "Can I hold him?" Sara asked. Billy handed Rupert to Sara, who was sitting on the floor. Rupert ran down her leg and jumped onto the floor. "Quick! Get him!" Billy called. Luckily Rupert was running towards Jimmy, who caught him. "Sorry," Sara said. "From now on I will remember that he is the fastest of us all!"

3. Helen's lesson

Today the Caterpillar Club was at Sara's house in the garden. They were watching her big sister Helen practice gymnastics. She could do some impressive flips and cartwheels. "I bet I could do that too!" Billy declared, always ready for an antic. He tried to do a cartwheel, but landed on his bum. Helen came over to him. "It is quite a silly antic to do gymnastics without any training. You could hurt yourself," she said sternly. "I wasn't trying to do an antic," Billy said. "I want to do a proper stunt!" "I can teach you how to do a pyramid," Helen suggested. Helen, Jimmy and Billy went on all fours next to each other. Rachel and Sara climbed on top of their backs. "We need another person to go on top of Rachel and Sara,"
Helen explained. "That sounds like a dangerous antic to me," Rachel objected. "Not if you do it properly," Helen replied. "I'll teach you next time!"

4. A famous portrait

Today the Caterpillar Club was meeting at the playground. Everyone was there apart from Rachel. Billy and Sara were on the swings. Jimmy was sitting on top of the slide keeping a look out for the Gloopy Goo Gang. Finally Rachel arrived, running towards them. "Look what I've got!" Rachel called out to the others. A round silver medal was hanging from a ribbon around her neck. "Come and wield it at me so I can see," Jimmy asked. Rachel took the medal in her hand to wield it proudly. "Remember the art competition I told you about?" she asked, "I came second!" Everyone had a look at the medal Rachel was able to proudly wield at them. "Congratulations!" they cheered. "What did your picture look like?" Sara asked when Rachel stopped to wield the medal. "It was a portrait of the four of us. I drew the Caterpillar Club with headquarters in the background." "That means we are famous!" Billy said triumphantly.

5. Tent or tree house?

Today the Caterpillar Club went to the Gloopy Goo Gang's headquarters, the tree house in Harry's garden. They wanted to see what their enemies were up to. When they arrived, they saw a large tent next to the tree. The Gloopy Goo Gang was sitting in front of it. "What are you doing with the tent?" Jimmy asked. "Are these your new headquarters?" Billy yelled. "This is only a temporary abode," Harry explained. "We are painting the tree house from the inside and outside." "Until the paint is dry, the tent will be our abode," Ollie added. "That's not a very secure abode," Sara pointed out. "Someone could easily break in." "Are you threatening to attack us?" Harry asked and jumped on his feet. "If I was planning an attack, I wouldn't tell you about it," Sara replied and laughed. "I was just commenting on your new abode." The Caterpillar Club left the Gloopy Goo Gang to finish painting their tree house.

6. As good as new

Today Jimmy arrived late at the playground. "I have just been to see the doctor," he explained. "Look! I no longer need a plaster on the cut on my arm. "The Caterpillar Club gathered around to inspect Jimmy's hand. The cut had healed well, but you could see a scar. "Wow," gasped Billy. "You can see exactly where the doctor had to suture the cut. I can see the marks left by the stitches. "Yes, when the doctor had to suture the cut, he made three stitches," Jimmy replied. "My mum has a similar scar," Rachel said. "She had a small operation on her tummy. You can see where they had to suture her tummy." "Same with my dad!" Billy exclaimed. "He has a scar on his leg from when he had an operation on his knee." "At some point in their life, most people will probably have to let a doctor suture something," Sara pondered.

7. The old shack

Today the Caterpillar Club was exploring an old, wooden shack in a field. The windows were broken. “I’m going to have a look around the back!” Billy yelled and ran off. The others had a closer look at the front door. It was nailed shut. Suddenly they heard a freaky noise warble through the house. “Maybe it’s ghosts who warble,” Sara whispered. “It’s coming from that window,” Rachel said. Jimmy had a look through it. It was Billy. He stopped making noises and laughed. “How did you get in there?” Jimmy called through the window. ”There was an open door at the back,” Billy yelled back. Jimmy, Rachel and Sara ran to the back and joined Billy. Everything inside was covered in dirt and cobwebs. “Were you scared when you heard me warble?" Billy asked. “Of course not," Sara insisted, "I knew only you would warble in such a silly way!"

8. Sara the scientist

Today Sara had invited the Caterpillar Club over to her kitchen. “I want to try an experiment with you!" said Sara. She put down several glasses in front of them. They were filled with colourful liquids blue, yellow, and red. "These are different drinks. I want you to tell me what they taste of," Sara explained. "I am trying to test your palate." Everyone took a sip of the yellow drink to put their palate to the test. "This is nice!" Rachel said. "It tastes like a fizzy orange drink." Next they tried they red one and everyone agreed it tasted like cranberry juice. Finally they tried the blue one. "It reminds me of a raspberry slushie!" Jimmy said. "It has the same colour." "I tricked your palate!" Sara said triumphantly. "They are all exactly the same lemonade. I just put food colouring in them. The colour tricks your palate into thinking that it tastes of what it looks like!" Everyone was impressed.

9. Mary's bike

Today Sara was late arriving at the Caterpillar Club headquarters. "I just got off the phone with our friend Mary. She was in bitter dudgeon. Someone stole her bike!" "What a nasty thing to happen," Rachel said. "I would be in dudgeon too." "Did she tell the police?" Jimmy asked. "I don't know," Sara replied. "I didn't get a chance to ask her. She was in such high dudgeon." "Sometimes the police catch a group of bike thieves and return all the bikes to their owners," said Billy. "I hope they manage to get Mary's bike back. She got it for her last birthday and really loved it. That's why she was in such bitter dudgeon about it," Sara explained. "We should keep our eyes open," Jimmy suggested. "Maybe we will see someone riding it. Then we can tell the police to arrest them." "But what if someone simply has the same bike as Mary?" Rachel wondered. Jimmy shrugged.

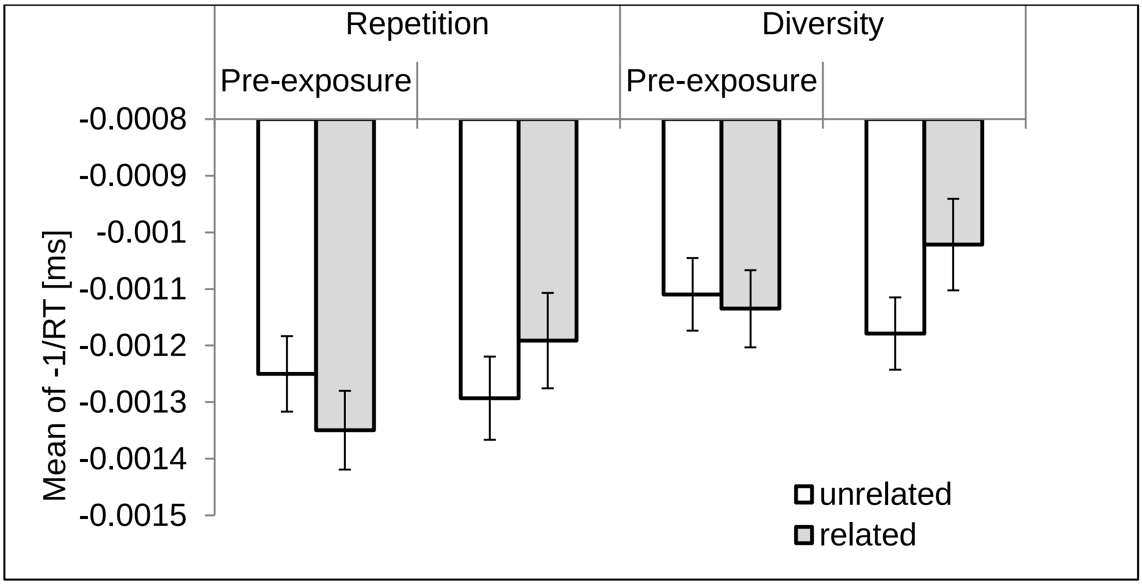
10. Mighty moles

Today the Caterpillar Club was playing ball games in Jimmy's garden. Jimmy threw the ball far over Billy's head. Billy tried to catch it, but stumbled backwards over a heap of earth on the grass. He got back up and brushed the grass off his trousers. "What is this heap doing here?" he said, sounding annoyed. "That's a mole hill," Jimmy explained. "Moles undermine our garden with their tunnels. Look, there are some more mole hills over there." "They are very hardworking animals," Sara added. "They use their strong paws as shovels to undermine the grass." "My dad really doesn't like them," said Jimmy. "He wishes they would stop trying to undermine the lawn. He has to mow the lawn around the mole hills." "I wish I could see a mole in action. It must look really cute when they undermine the earth," Rachel wondered. "Unfortunately, we are too big to go down into their tunnels," Sara said. They all laughed.

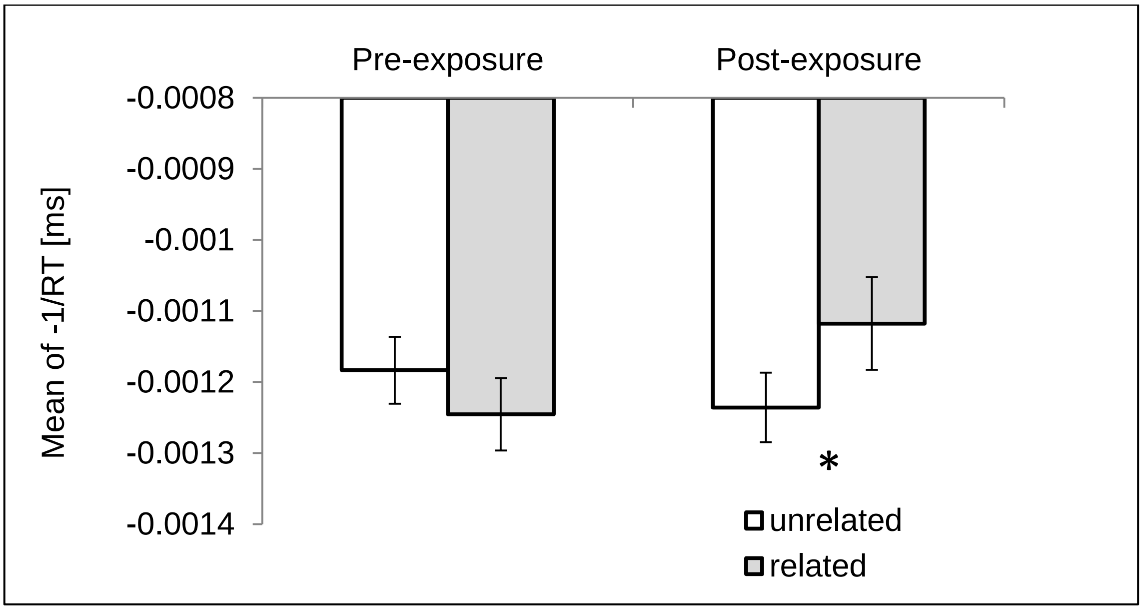
Appendix 5.2

Inverse transformed data for masked priming lexical decision

5.2.1 Data split by exposure condition. Error bars show 95% confidence intervals.



5.2.2 Data collapsed across exposure conditions. Error bars show 95% confidence intervals.



Appendix 5.3

Notes. 'Condition' refers to repetition vs. diversity.

5.3.1 Reading accuracy during exposure sessions

Formula: Accuracy ~ TOWRE + Condition + (1 | Subject) + (Condition | Item) (Family = binomial)

Fixed effects:

	Estimate	***SE***	***z***	***p***
Intercept	-0.65	1.35	-0.48	.631
TOWRE	0.03	0.01	2.26	.024
Condition	-0.12	0.46	-0.27	.790

5.3.2 Orthographic decision

Formula: Accuracy ~ TOWRE + Condition + (1 | Subject) + (Condition | Item) (Family = binomial)

Fixed effects:

	Estimate	***SE***	***z value***	***p***
Intercept	0.08	2.64	0.03	.975
TOWRE	0.02	0.02	0.96	.335
Condition	-0.71	0.89	-0.79	.428

5.3.3 Spelling

Notes. The maximal model did not converge; to simplify the model, the high correlation (1.0) between the random intercepts for item and the by-item random slopes for condition was removed. The difference in explained variance between the maximal model and the simplified final model was not significant.

Formula: Accuracy ~ TOWRE + Condition + (1 | Subject) + (0 + Condition | Item) + (1| Item) (Family = binomial)

Fixed effects:

	Estimate	***SE***	***z value***	***p***
Intercept	-1.66	1.75	-0.95	.345
TOWRE	0.03	0.02	2.08	.038
Condition	-0.56	0.53	-1.06	.287

5.3.4 Definition knowledge

Notes. The maximal model did not converge; to simplify the model, the high correlation (.99) between the random intercepts for item and the by-item random slopes for condition was removed. The difference in explained variance between the maximal model and the simplified final model was not significant.

Formula: Accuracy ~ TOWRE + Condition + (1 | Subject) + (0 + Condition | Item) + (1| Item) (Family = binomial)

Fixed effects:

	Estimate	***SE***	***z value***	***p***
Intercept	-3.16	1.70	-1.86	.063
TOWRE	0.03	0.01	2.37	.018
Condition	-1.27	0.45	-2.82	.005

5.3.5 Masked priming

5.3.5.1 Main model

Formula: (-1/RT) ~ TOWRE + Condition * Session * Relationship + (Session * Relationship | Subject) + (Condition * Session * Relationship | Item)

Fixed effects:

	Estimate *(10^{-6})*	***SE*** *(10^{-6})*	***t***
Intercept	-1186.00	161.70	-7.33
TOWRE	-.08	1.33	-.06
Condition	167.00	93.38	1.79
Session	28.34	32.62	0.87
Prime-Target Relationship	17.59	33.17	0.53
Relationship x Session	173.70	56.40	3.08
Relationship x Condition	56.69	62.20	0.91
Session x Condition	-24.51	64.03	-0.38
Condition x Session x Relationship	-9.56	109.60	-0.09

5.3.5.2 Pairwise comparisons to explore prime-target relationship x test session interaction

Pre exposure only:

Formula: (-1/RT) ~ TOWRE + Relationship + (Relationship | Subject) + (Relationship | Item)

Fixed effects:

	Estimate *(10^{-6})*	***SE*** *(10^{-6})*	***t***
Intercept	-1404	257.5	-5.45
TOWRE	1.75	2.23	0.78
Relationship	-68.58	35.11	-1.95

Post exposure only:

Formula: (-1/RT) ~ TOWRE + Relationship + (Relationship | Subject) + (Relationship | Item)

Fixed effects:

	Estimate *(10^{-6})*	***SE*** *(10^{-6})*	***t***
Intercept	-1068	299.4	-3.57
TOWRE	-0.96	2.50	-0.39
Relationship	108.1	48.32	2.24

Appendix 6

6.l Combined analysis of pre-exposure masked priming lexical decision

Formula: (-1/RT) ~ TOWRE + Relationship + (Relationship | Subject) + (Relationship | Item)

Fixed effects:

	Estimate *(10^{-6})*	***SE*** *(10^{-6})*	***t***
Intercept	-761.20	163.40	-4.66
TOWRE	-2.70	1.42	-1.90
Relationship	-43.28	13.44	-3.22

6.2 Combined analysis of post-exposure masked priming lexical decision

Formula: (-1/RT) ~ TOWRE + Relationship + (Relationship | Subject) + (Relationship | Item) + (Relationship | Condition)

Fixed effects:

	Estimate *(10^{-6})*	***SE*** *(10^{-6})*	***t***
Intercept	-357.20	164.00	-2.18
TOWRE	-6.37	1.41	-4.52
Relationship	-75.47	24.30	-3.13