

*Becoming a written word: eye movements
reveal order of acquisition effects
following incidental exposure to new
words during silent reading*

Article

Accepted Version

Joseph, H. S. S. L., Wonnacott, E., Forbes, P. and 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). pp. 238-248.
ISSN 0010-0277 doi:
<https://doi.org/10.1016/j.cognition.2014.06.015> Available at
<http://centaur.reading.ac.uk/42501/>

It is advisable to refer to the publisher's version if you intend to cite from the
work. See [Guidance on citing](#).

Published version at: <http://www.sciencedirect.com/science/article/pii/S0010027714001255>

To link to this article DOI: <http://dx.doi.org/10.1016/j.cognition.2014.06.015>

Publisher: Elsevier

All outputs in CentAUR are protected by Intellectual Property Rights law, including copyright law. Copyright and IPR is retained by the creators or other copyright holders. Terms and conditions for use of this material are defined in the [End User Agreement](#).

www.reading.ac.uk/centaur

CentAUR

Central Archive at the University of Reading

Reading's research outputs online

Becoming a written word: Eye movements reveal order of acquisition effects following
incidental exposure to new words during silent reading

Holly S. S. L. Joseph*^{1,2}, Elizabeth Wonnacott³, Paul Forbes² & Kate Nation²

¹ Oxford Brookes University, Gipsy Lane, Oxford, OX3 0BP

² University of Oxford, South Parks Road, Oxford, OX1 3UD, UK

³ University of Warwick, University Road, Coventry, CV4 7AL

*Corresponding author

Department of Psychology, Social Work and Public Health

Oxford Brookes University

Gipsy Lane Campus

Headington

Oxford

OX3 0BP

United Kingdom

E-mail: hjoseph@brookes.ac.uk

Telephone: 01865 483786

Abstract

We know that from mid-childhood onwards most new words are learned implicitly via reading; however, most word learning studies have taught novel items explicitly. We examined incidental word learning during reading by focusing on the well-documented finding that words which are acquired early in life are processed more quickly than those acquired later. Novel words were embedded in meaningful sentences and were presented to adult readers early (day 1) or later (day 2) during a five-day exposure phase. At test adults read the novel words in semantically neutral sentences. Participants' eye movements were monitored throughout exposure and test. Adults also completed a surprise memory test in which they had to match each novel word with its definition. Results showed a decrease in reading times for all novel words over exposure, and significantly longer total reading times at test for early than late novel words. Early-presented novel words were also remembered better in the offline test. Our results show that order of presentation influences processing time early in the course of acquiring a new word, consistent with partial and incremental growth in knowledge occurring as a function of an individual's experience with each word.

Key words: word learning reading Age of Acquisition eye movements

1. Introduction

Once children have grasped the principle that letters code for speech sounds, at least in an alphabetic orthography such as English, they have in place the tools available to build orthographic knowledge via their reading experience. Early in reading development it is likely that children are reading words that they are already familiar with in the oral domain – they are learning to map orthography onto their existing phonological and semantic knowledge. At the same time however, we know that children learn new words from their reading experience and that from mid-childhood onwards, most new vocabulary is learned via reading (e.g., Anderson & Freebody, 1981; Landauer & Dumais, 1997; Nagy, Anderson & Herman, 1987; Nagy, Herman, & Anderson, 1985). It follows from this that children must regularly see novel words in print, yet we know relatively little about how children learn new words in this situation, or the parameters that influence learning. In this paper, we examine incidental word learning during reading by focusing on an effect which has been reported across many studies: that words which are acquired early in life are processed more quickly than those acquired later.

1.1 Orthographic learning in children and adults

The self-teaching hypothesis (Share, 1995) sets out how children acquire orthographic knowledge. A growing literature supports the central tenants of this hypothesis, namely that phonological decoding is the fundamental basis of orthographic learning and that each decoding attempt provides the child with a word-specific encounter, as well as adding to their accumulated knowledge base about how orthography works in their language (see Share, 2008 for a review). Experiments conducted within the self-teaching framework have considered a number of questions including the number of exposures children need to subsequently recognise newly learned words and discriminate them from a phonological or visual foil and the durability of learning over time (e.g., Nation, Angell & Castles, 2007; Share, 2004), whether learning is influenced by words being presented or tested in

context or in isolation (e.g., Landi, Perfetti, Bolger, Dunlap & Foorman, 2006; Martin-Chang, Levy, & O'Neil, 2007; Nation et al., 2007; Wang, Castles, Nickels & Nation, 2011), whether knowledge of the meaning of the phonological form of new words prior to orthographic exposure influences ease of learning (Wang et al., 2011; Ouellette & Fraser, 2009) and the extent to which learning is modulated by orthographic consistency (e.g., Wang et al., 2011; Wang, Castles & Nickels, 2012). Summarizing a complex and sometimes contradictory literature, there is evidence that children learn new orthographic forms relatively easily from minimal incidental exposure, at least when tested using recognition paradigms, and this knowledge is retained over time. For example, Share (2004) found that children learning to read Hebrew showed no decline in learning when re-tested one month after exposure. Nation et al. (2007) reported a different pattern in children learning to read in English: although retention declined when the children were retested one week after exposure, significant learning was still evident. Knowledge of the meaning of the to-be-learned words facilitates learning, as does presenting the words in context, but only when the words are irregular in terms of spelling-sound mappings (Wang et al., 2011; Wang, Nickels, Nation & Castles, 2013).

One issue with the orthographic learning via self-teaching literature concerns how to measure learning. Generally, two types of metric are employed: recognition paradigms that are variants of orthographic choice in which children have to discriminate learned items from foils, and recall paradigms in which they attempt to spell the words. It is not unusual for these two metrics to provide a different estimate of learning, even within the same experiment. Learning is typically better when assessed via recognition paradigms. Arguably however, these are not sensitive as performance tends to be high. Assessing learning via spelling also tends towards lack of sensitivity as performance levels tend to be quite low, presumably because recalling a spelling requires a greater specification of its orthographic form than recognising it does (Wang et al., 2011). In addition, although the learning of orthographic forms in Share's earlier experiments was a consequence of incidental exposure, in line with the notion of self-teaching, efforts to induce meaning to the target words has resulted in paradigms

that contain a hefty dose of explicit teaching. For example, Wang et al. (2011) pre-exposed children to the meaning of the phonological forms of words by showing them pictures that illustrated each word's referent and function. Children were explicitly instructed to learn the meanings and this was tested multiple times, before the orthography of the new words was presented. This is quite different to children's naturalistic reading experiences in two ways. First, children are rarely taught and tested on new vocabulary in this explicit and direct way and second, knowledge of vocabulary doesn't always pre-date orthographic exposure: many new words are initially encountered in text, at least by mid-childhood. It is also worth noting that only a small number of words can be trained in this explicit way as children find this type of learning difficult (e.g., Ricketts, Bishop, Pimperton, & Nation, 2011).

Parallel to these studies, experiments with adults have explored the acquisition of new orthographic forms from a different theoretical perspective. Bowers, Davis and Hanley (2005) exposed adults to new forms (e.g., *banara*) via a repeated typing exercise and then examined the effect this had on the processing of existing words (e.g., *banana*). Introducing a new neighbour resulted in slower semantic classification times for the existing base word, consistent with the new word inducing lexical competition. Qiao, Forster and Witzel (2009) questioned whether Bowers et al. provided convincing evidence of lexicalisation, arguing that their data could be explained via an episodic learning account. Subsequent studies have shown that if learners are exposed to new orthographic forms in the context of quite extensive and explicit semantic training, lexicalisation does occur, as evidenced by the emergence of a prime lexicality effect (Qiao & Forster, 2012) and reliable semantic priming under masked conditions (Tamminen & Gaskell, 2013).

It is important to note the repetitive and explicit nature of the training regimes employed in these studies with adults, with many exposures to the new orthographic forms and multiple sessions needed to induce meaning to the new forms using tasks such as word-picture matching, picture-word matching and word-definition matching. These studies with adults demonstrate the interactive and dynamic nature of lexical learning, but they fail to touch base with the developmental literature, though

studies have begun to explore these issues for *spoken* words, e.g., Brown, Weighall, Henderson & Gaskell, 2012; Henderson, Weighall, Brown & Gaskell, 2013). A small number of studies have examined *written* word learning in adults from a developmental perspective, attempting to mimic the learning conditions facing a child. These though have focused on the influence of orthographic consistency and regularity on reading single words aloud, and have used intensive training regimes with explicit instruction and feedback (e.g., McKay, Davis, Savage & Castles, 2008; Taylor, Plunkett & Nation, 2011).

1.2 Age-of-acquisition and lexical processing

It is well established that the age at which new words are acquired during childhood influences how quickly they are processed in adulthood (for reviews, see Brysbaert & Ghyselinck, 2006; Johnston & Barry, 2006; Juhasz, 2005). Age of acquisition (AoA) accounts for unique variance in word naming and lexical decision in adults, even after a broad range of psycholinguistic variables are accounted for (e.g., Cortese & Khanna, 2007; Morrison & Ellis, 1995). AoA effects have also been observed in children's reading behaviour. Coltheart, Laxon and Keating (1998) constructed lists of early vs. late acquired words, matched for frequency and imageability, with AoA estimated as the age at which items entered a child's oral vocabulary; 9-year-olds read early acquired words more accurately than later acquired words. Defining AoA as the age at which children first experienced a word in print, Nazir, Decoppet, and Aghababian (2003) reported that this metric influenced visual lexical decision performance in children from 1st-to-5th grade.

Two theoretical accounts of AoA enjoy support in the literature, the semantic hypothesis (Brysbaert, Wijnendaele, & Deyne, 2000; Ghyselinck, Lewis & Brysbaert, 2004), which proposes that AoA is an intrinsic property of the semantic representation of words (and objects), and that the magnitude of AoA effects observed is associated with the amount of semantic involvement that is required in a given task, and the mapping/loss of plasticity hypothesis (Ellis & Lambon Ralph, 2000),

which predicts an advantage for early-learned words because they cause larger changes in connection weights than those learned later, resulting in a loss of plasticity in the network and the ensuing decrease in efficiency for learning later words. Our intention here is not to debate the basis of AoA effects, but we note that the effect is well-replicated across different lexical tasks and languages, and that it is not completely attributable to other variables including frequency (e.g., Bonin, Barry, Meot, & Charlard, 2004; Ghyselinck, Lewis & Brysbaert, 2004) and imageability (e.g. Monaghan & Ellis, 2002).

An important insight from efforts to model AoA computationally is that order-of-acquisition (OoA) effects emerge as natural property of a system that learns over time (Steyvers & Tenenbaum, 2005). Monaghan and Ellis (2010; see also Ellis & Lambon Ralph, 2000) modelled OoA effects on word naming, using Harm and Seidenberg's (1999) connectionist model of word reading development. When stimuli are encountered cumulatively and in an interleaved fashion, and when input-output mappings have some degree of inconsistency (as the mappings between orthography and phonology do in English), both frequency and order of learning are significant factors influencing the model's performance. Monaghan and Ellis note that when input-output mappings are more inconsistent and unpredictable, greater effects of OoA are to be expected. This is consistent with empirical findings that when a behavioural task taps the arbitrary mappings between form and semantics (for example, object naming and lexical decision), AoA effects are greater than for tasks such as reading aloud which tap more systematic mappings between phonology and orthography (e.g., Johnston & Barry, 2006; Juhasz, 2005). Importantly however, Monaghan and Ellis' model demonstrates that semantic knowledge per se is not critical for age or order effects to emerge, as their model did not include semantics.

If OoA is a natural property of a learning system, we should see evidence of order effects when we ask learners to acquire new material under experimental conditions. Stewart and Ellis (2008) showed exactly this, finding that categorization of novel checkerboard patterns was influenced by the order in which stimuli were presented during training. Within the oral language domain, as people learn new phonological labels for familiar items (e.g., English speakers learning the Spanish labels),

order effects are seen such that words introduced early in training enjoy a processing advantage relative to words encountered later in training, despite overall frequency being matched (Izura, Perez, Agallou, Wright, Martin, et al., 2011; see also Tamminen & Gaskell, 2008). Catling, Dent, Preece and Johnston (2013) created a laboratory analogue of AoA in object naming by training participants to associate novel phonological forms with Greeble stimuli. Mirroring the AoA effects seen in object naming, OoA effects were observed such that Greebles experienced early in training were recognised more readily and named more quickly than Greebles experienced later in training. Taken together, these laboratory training studies of OoA provide support for the view that AoA effects are a property of general learning systems; arguably however all are limited in that they train language in an artificial way. For example, Izura et al. (2011) used a paired associate learning paradigm in which participants viewed a picture of a (familiar) object while hearing its spoken form in an unfamiliar foreign language; they were then instructed to repeat and remember the words. Although clear effects of OoA were observed, it is not clear whether this would maintain without the explicit instruction to learn. There are also potential concerns about transfer effects from native language, given people were learning new names for familiar objects. Finally, words were learned by viewing pictures and hearing their spoken form. Vocabulary is rarely acquired in this way, with the vast majority of new vocabulary being learned incidentally via reading from mid-childhood onwards.

In summary, there is evidence that OoA effects emerge in tasks where the learner is explicitly learning new mappings – i.e. between checkerboards and categories, between novel words and existing words and between new words and new pictures. The suggestion is that the same learning processes underlie the AoA effects seen in lexical processing. Clearly however, very little of our vocabulary learning, whether via oral communication or via text, occurs in this explicit manner. Experiments have not considered order effects following more naturalistic exposure, and in particular, none have examined order effects in learning during reading.

1.3 Eye movements, AoA and reading novel words

Much has been learned about reading from monitoring people's eye movements as they read text (for review, Rayner, 1998; 2009). As well as providing a moment-by-moment measure of reading behaviour, no secondary task or phonological output is required (cf. lexical decision or reading aloud). It thus offers an ideal methodology to study reading processes under more naturalistic conditions. Juhasz and Rayner (2006) conducted a carefully controlled eye movement experiment in which they embedded target words manipulated for AoA but controlled for frequency (and for familiarity, length, concreteness and imageability in subsets of items) in sentences. Results showed effects of AoA early in processing, with the very first fixation on late-acquired words being longer than that on early-acquired words; this effect was consistent across later processing measures too. Strikingly, these findings show that during normal sentence reading and with no explicit instruction or task demands, the age at which you first acquired a word has a strong and immediate effect on how long you look at it.

To our knowledge, eye movement methodology has not been employed in studies of orthographic learning in children or adults, although a small number of experiments have examined how readers respond to a novel word embedded in an otherwise meaningful text. Consistent with the well-replicated effect of longer first pass fixation durations on low than high frequency words (e.g., Inhoff & Rayner, 1986; Rayner & Duffy, 1986), adults show significantly longer first fixations on novel words (Chaffin, Morris & Seely, 2001; Lowell & Morris, 2014), showing that novel words are treated like potential or existing words on their first encounter. Readers also make more regressions back to novel words and spend longer re-reading them (Chaffin et al., 2001), suggesting that readers use subsequent contextual information to infer a possible meaning for a new form. Finally, readers show larger length effects on novel than familiar words (Lowell & Morris, 2014), reflecting efforts to encode the new letter strings, with long words placing greater demands on processing than short words. In summary, previous research suggests that on a single exposure to a novel word, skilled readers are encoding its form, and attempting to infer meaning by integrating it with subsequent context. We still

do not know, however, what happens when readers encounter the novel word on subsequent occasions. With multiple exposures, new words increase in frequency and are no longer novel. It is reasonable to predict that reading times would become shorter, as words become more familiar. If, however, OoA is also relevant, reading times should not only be predicted by quantity of exposure, but also by how early in learning a word is experienced. This is the basic premise of the experiment we describe in this paper.

1.4 Our Experiment

Current methods to investigate orthographic learning are limited. Although incidental learning was tapped by Share's original self-teaching paradigm, measuring learning in that paradigm is problematic. Experiments that probe the learning of new forms by measuring competition effects on existing words (e.g., Bowers et al., 2005; Qiao & Forster, 2012) offer a promising way forward, but all involve explicit training regimes that are far removed from the situation we (or children) face as we encounter a novel word in text in our normal reading experiences. As reading is the primary vehicle by which new vocabulary is acquired from mid-childhood onwards, it is important to understand more about how new words are processed under circumstances that mimic natural reading experiences as closely as possible, i.e., silent reading of meaningful text with no secondary task. In our experiment, we embedded novel words into sentences and monitored eye movements as adults read them silently and with no instruction to learn or remember the words. Words were presented in different meaningful sentences over a number of days, allowing us to examine changes in eye movements to the words over time, and at a final test session. Our key manipulation was OoA. All words were presented 15 times, but some words were presented early (first seen on day 1) and others late (first seen on day 2). This provided a laboratory analogue of AoA, and allowed us to examine the effect of order on the learning of new words via incidental exposure during reading. Learning was tested in two ways: implicitly, by measuring reading times on the target words embedded in neutral sentences and via a surprise memory task which assessed explicit learning by asking participants to match the newly learned

words with their semantic properties. We predicted that during the exposure phase, reading times would decrease over time in both the early and the late condition as a function of exposure, but that the slope would be steeper for early items. During test, we predicted that participants would exhibit shorter reading times in the early condition and that they would also remember the meanings of more early-presented words than late ones in the surprise memory test. Finally, as the newly learned words were embedded in neutral sentences at test, we anticipated that reading times would be longer, relative to when the same words were embedded in meaningful sentences at the end of the exposure phase. This would be due to increased difficulty in accessing novel word meanings without contextual cues, especially for late-presented items.

2. Method

2.1 Design

The exposure phase of the experiment had a 2 (OoA: early vs. late) x 5 (day: 1-5) within-participant, within-item design. All participants saw half the novel words in the early condition and half in the late condition. Learning was assessed after the exposure via reading times of novel words in neutral sentences, and via an explicit memory test.

2.2 Participants

Data were collected from 41 participants, all of whom were students and staff members at the University of Oxford. Participants were given a chocolate bar every day of testing, or paid £20 when they had completed all five days of testing. Eye movement data from four participants were excluded due to tracker loss or technical difficulties leaving 37 complete data sets for analysis.

2.3 Materials

2.3.1 To-be-learned words

Sixteen nonwords were constructed by randomly generating 32 consonant-vowel-consonant (CVC) strings and then randomly pairing the strings to form 16 bisyllabic six letter nonwords. We then

excluded the following: (1) any CVCs that formed actual words; (b) any CVCs containing Q, X, or J (to minimize pronunciation difficulties) and (c) any six-letter nonwords with a double consonant in the middle. We replaced excluded items by generating more CVC strings and concatenating them. This resulted in the following nonwords which were arbitrarily assigned to one of two lists of eight words; List 1: wefdof, hebnin, lisbaw, lamcuk, sorlas, leskud, zomlal, sabrup; List 2: gadwob, gogruc, felror, helvof, musdit, guvpaf, hosfot, lurcor. The assignment of these lists as “early” or “late” words was counterbalanced across participants: Group A had List 1 as their early words and List 2 as their late words, and vice versa for Group B.

Each nonword was assigned a meaning, which related to a tool used by an imaginary tribe. This meaning was conveyed by embedding the word in a set of 15 sentence contexts (one set of 15 sentences per nonword) which were presented to participants as separate trials in the experiment. Each context provided (fictional) semantic information about various aspects of the tool, for example, ‘It can be said that a lurcor is a type of cooking pot’ (see Appendix A for an example of a full set of sentences). Note that the assignment of each nonword to a specific set of 15 sentences was held constant across the counterbalancing – i.e., the novel word *lurcor* shown in Appendix A occurred with these specific sentences both when it was assigned as a late learned word (for Group 1 participants) and an early learned word (for Group 2 participants). This counterbalancing ensured that any overall differences between early and late learned words could not be due to differences in phonological form or the semantic contexts in which they were learned. Target words were never presented as the first or the last word in a sentence.

We also created a set of 16 test sentences, also shown in Appendix A. In contrast to the exposure sentences, care was taken to make the contexts as neutral as possible so as not to provide any specific clues as to the meaning of the word.

2.3.2 Exposure Schedule.

Each of the 16 target words appeared in 15 different sentences during the exposure phase. Each sentence was read once and the sentences were distributed over five sessions, with one session per day, Monday-Friday across a week. Only early words were seen on day 1. On days 2-4, more late than early words were presented so that by day 5, early and late items had an identical cumulative frequency of 15 (see Table 1). Whether the early or late items were presented first during each session was alternated. At test, items were presented in a random order.

Table 1: Exposure schedule for early and late items and their contexts

Session/Day	Early Items		Late Items	
	N exposures	Sentence ID	N exposures	Sentence ID
Day 1	6	1-6	0	-
Day 2	2	7-8	4	1-4
Day 3	2	9-10	4	5-8
Day 4	2	11-12	4	9-12
Day 5	3	13-15	3	13-15
Test (Day 5)	1	Neutral	1	Neutral

Note: Sentence ID refers to the context provided by each sentence, see Appendix A.

2.4 Apparatus.

Throughout each exposure session and at test, participants' eye movements were recorded using a desktop Eyelink 1000 eye tracker (SR Research; Mississauga, Canada) as they read sentences from a 14" computer monitor at a viewing distance of 60cm, and each character covered 0.24° of horizontal visual angle. Sentences were presented in white, Courier New font size 15, on a black background. Eye movements were monitored at a rate of 1000Hz. Although participants read binocularly, only the movements of the right eye were monitored. Participants leaned on chin and forehead rests during the experiment to minimise head movements.

2.5 Procedure

Participants were tested on five consecutive days for approximately 10 minutes each day. Before the start of the experiment participants were told that they would read silently a series of sentences about some tribal tools. They were informed that some of the words might be unfamiliar and were instructed to read the sentences carefully but naturally at their normal reading speed.

During each exposure session, an initial calibration of the eye tracker was carried out during which the participant was asked to look at each of three fixation points, extending in a horizontal line across the centre of the screen, while their fixation position was recorded for each point. Once the eye tracker had been calibrated with satisfactory accuracy, three practice trials (which repeated the verbal instructions given by the experimenter) were presented immediately followed by the experimental sentences. Each sentence was preceded by the appearance of a small fixation box. Participants were instructed to fixate the box in order to trigger the appearance of the sentence, thus ensuring accuracy for each trial; after reading, they pressed a button to terminate the trial. Following each sentence, calibration accuracy was checked and the tracker was recalibrated if necessary.

On day 5, immediately after the final exposure, participants read silently 16 sentences containing each of the target words in a neutral context. They were not given any additional instructions so that from a participant perspective this “test” session was not distinguishable from the training sessions. The order of the sentences was randomised for each participant. Following this, there was a surprise memory task in which participants were asked to match each of the 16 (written) tool names to one of 16 written descriptions of its function (e.g., ‘lurcor’ would be matched to ‘cooking pot’; see Appendix B).

3. Results

3.1 Performance on the surprise memory test

Performance levels were low. Out of a possible 16, mean matching performance for the early and late items was 4.5 (SD = 2.1) and 3.0 (SD = 1.6) respectively. A Monte Carlo simulation,

summarised in Table 2, showed that the probability of a participant obtaining a score of four or more out of 16 by chance was less than .05. Thus, participant scores were reliably above chance in the early condition only. Although performance levels were low, a within-item paired t-test showed that more early items were correctly matched than late ones, $t(15) = 2.35$, $p = .03$, in line with our prediction that early acquired words would be learned more easily than late acquired words. (Note that we were unable to conduct analyses by-participant as the within-participant scores for Early and Late would not be independent: getting any item correct affected the probability of that participant getting any another item correct.)

Table 2: Probability (computed via Monte-Carlo simulation) of obtaining each score by chance in the surprise memory test which involved pairing each of 16 tool words to the correct description of its function.

Number of correct matches	Probability
0	1.0000
1	0.6370
2	0.2716
3	0.0772
4	0.0185
5	0.0046
6	0.0008

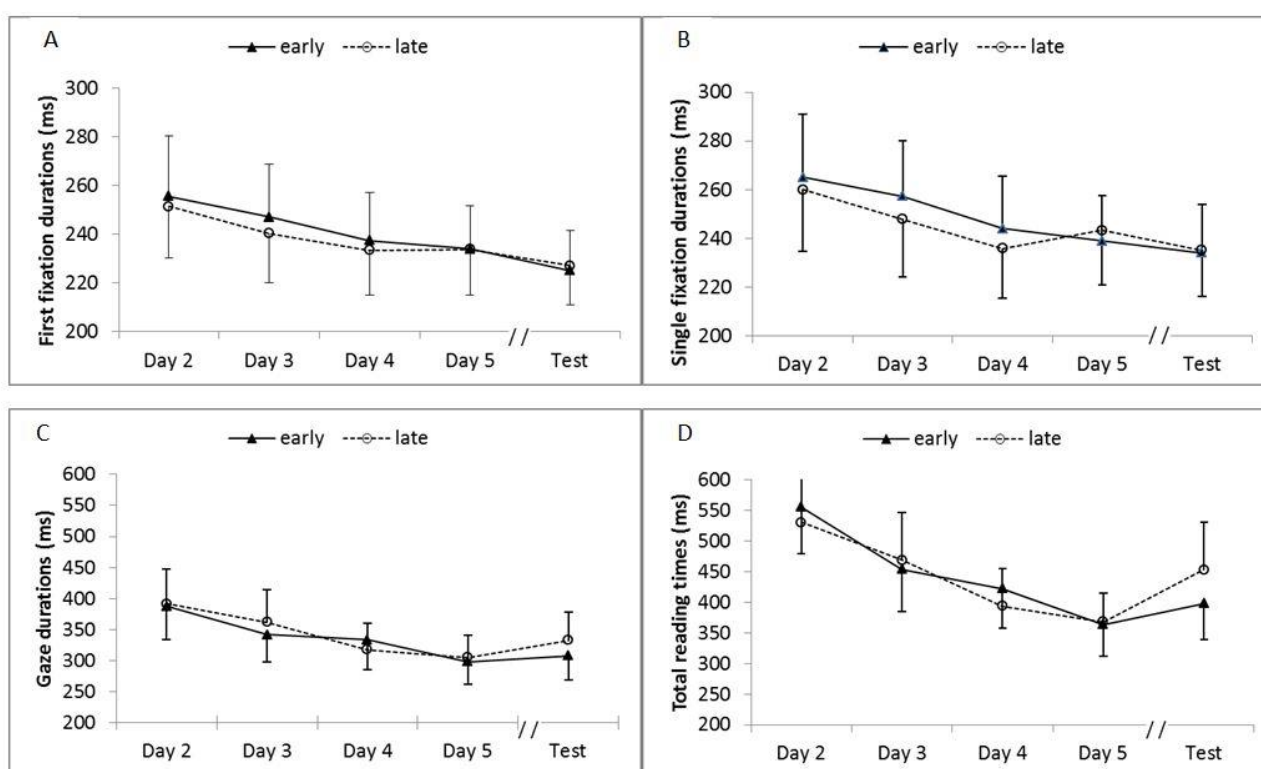
3.2 Analyses of eye-movements

We recorded eye-movements throughout the five exposure sessions and as participants read the neutral test sentences at the end of day 5. Data from exposure and test were analysed separately. For all analyses, fixations shorter than 80ms and longer than 1200ms were excluded and trials which showed blinks or tracker loss on the target word were also deleted. This resulted in the loss of less than 1% of the eye movement data in total, and the amount of data excluded did not differ between early and late conditions in any of the measures.

3.2.1 Eye movement data in the exposure phase

Figure 1 shows mean reading times for the four eye movement measures of interest across exposure: first fixation (the duration of the first fixation made on the target); single fixation duration (the duration of a fixation on the target, provided the word received only one fixation during first pass); gaze duration (the sum of all first pass fixations made on the target); and total reading time (the sum of all fixations made on the target, including second pass reading times). Note that data from day 1 are excluded from the analyses that follow since there were no late-presented words on that day.

Figure 1: Reading times during training and test sessions. Panel A: first fixation durations; Panel B: Single fixation durations; Panel C: Gaze durations; Panel D: Total reading times



Data were analysed in the R computing environment (R Development Core Team, 2010) using linear mixed models (Baayen, 2008; Jaeger, 2008; Quené & van den Bergh, 2008). We were interested in whether, for any of the measures, (i) reading times decreased with day of testing (i.e. a frequency of exposure effect) (ii) there was an overall effect of OoA (iii) the effect of OoA differed by day. To

explore this, we used the same methodology for each eye-movement measure. First, we ran a model with the eye movement measure as the predicted variable, fixed factors of OoA (binary: early vs. late) day (numeric: 2-5) and the interaction between OoA and Day, random intercepts for participants and items and random by-participant and by-item slopes for OoA, day and the interaction (i.e., a full random slopes structure; see Barr, Levy & Scheepers & Tilly, 2013¹). We then compared this full model to models with (i) the fixed effect of Day removed (ii) the fixed effect of OoA removed and (iii) the interaction removed.

For first fixations, model comparison showed an effect of Day ($\chi^2 = 20.5$, $df = 1$, $p < .001$; *Ms*: Day 2 = 253, Day 3 = 243, Day 4 = 235, Day 5 = 234) such that first fixation durations decreased over time. However, there was no reliable effect of OoA ($\chi^2 = 2.8$, $df = 1$, $p = .09$, *Ms*: early = 243; late = 240) and no interaction ($\chi^2 = 0.13$, $p = .72$).

For single fixations, there was an effect of Day ($\chi^2 = 10.3$, $df = 1$, $p < .005$, *Ms*: Day 2 = 262, Day 3 = 251, Day 4 = 238, Day 5 = 241) but no effect of OoA ($\chi^2 = 0.87$, $df = 1$, $p = .34$, *Ms*: early = 248; late = 246) and no interaction ($\chi^2 = 0.28$, $p = .60$).

For gaze durations, we saw the same pattern: an effect of Day ($\chi^2 = 35.56$, $df = 1$, $p < .001$; *Ms*: Day 2 = 391, Day 3 = 356, Day 4 = 318, Day 5 = 293), no effect of OoA ($\chi^2 = 0$, $df = 1$, $p = 1$; *Ms*: early = 332, late = 343) and no interaction ($\chi^2 = 0$, $df = 1$, $p = 1$).

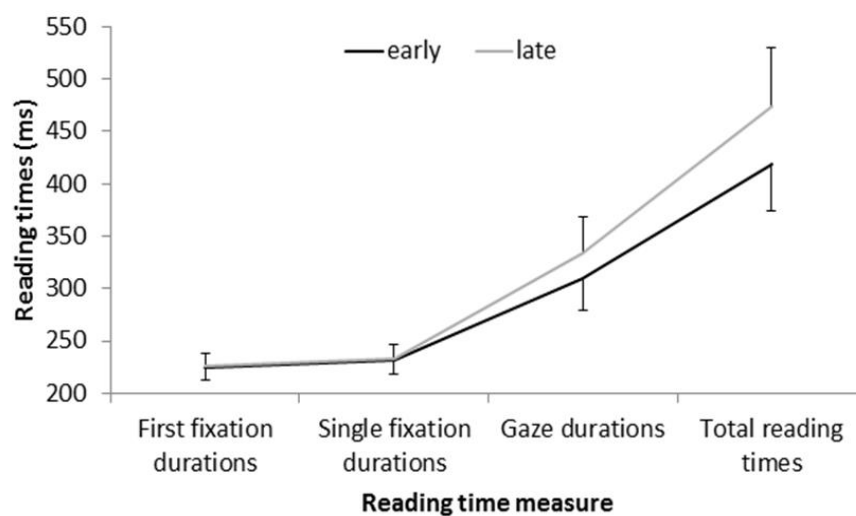
Finally, for total times, once again there was an effect of Day ($\chi^2 = 35.43$, $p < .001$, *Ms*: Day 2 = 539, Day 3 = 464, Day 4 = 404, Day 5 = 366), no effect of OoA ($\chi^2 = 1.03$, $df = 1$, $p = .31$; *Ms*: early = 441, late = 445), and no interaction ($\chi^2 = 0.16$, $df = 1$, $p = .69$). Thus, the same pattern was observed for each dependent measure, with reading times decreasing with increased exposure, but no significant difference for early vs. late acquired words and no evidence that this differed by day.

3.2.1 Eye movement data in the test phase

¹ For first fixations only, the full model did not converge and we had to remove the random by items slope for the interaction.

Next we examined reading times for the target words at test when they were embedded in neutral sentences. Data are summarised in Figure 2. Of interest is whether there was a main effect of OoA for any of the eye-movement measures. For each, we ran a model predicting that measure with a fixed effect of OoA, random intercepts for participants and items and random by-participant and by-item slopes for OoA. In each case we compared the full model with a model in which the fixed effect of OoA was removed. There was no reliable effect of OoA in first fixation duration ($\chi^2 = 0.48$, $df = 1$, $p = .49$, M_s : early = 221; late = 226) or single fixation duration ($\chi^2 = 0.52$, $df = 1$, $p = .47$, M_s : early = 231; late = 233); a hint of an effect is seen for gaze duration ($\chi^2 = 2.08$, $df = 1$, $p = .15$, M_s : early = 310; late = 333) and for total reading times, there was a significant effect of OoA ($\chi^2 = 5.52$, $df = 1$, $p < .05$, M_s : early = 419; late = 474) with longer reading times for late-presented words relative to those presented early.

Figure 2: Reading times for early vs. late words in neutral sentences (test)



The effect of OoA for the total time measure in the test sentences is particularly striking given that we did not see an equivalent effect of OoA during exposure. This is especially noteworthy when we compared test data with data from the final exposure session from day 5. Recall that this immediately preceded the presentation of the test sentences, and that participants were not aware that the test phase was any different to what they had experienced throughout exposure. The only difference was that at

test, the target words were presented in a neutral context whereas throughout exposure, the target words appeared in a meaningful context. Although there was no significant interaction between OoA and Day during the exposure phase, given the effect within the test data, we ran an equivalent analysis using day 5 exposure data only. This confirmed there to be no effect of OoA ($\chi^2 = 0.09$, $df = 1$, $p = .76$). Further to this, to test our prediction that reading times would be longer in the neutral than meaningful sentences, we compared test vs. exposure for day 5 only, using a model that included fixed factors of sentence type (meaningful at exposure vs. neutral at test), OoA, and their interaction plus random intercepts for participants and items and full random slopes structure. There were marginal effects of OoA ($\chi^2 = 4.91$, $df = 2$, $p = .086$) and sentence type ($\chi^2 = 4.95$, $df = 2$, $p = .08$), and a marginal interaction ($\chi^2 = 3.56$, $df = 1$, $p = .059$). Breaking the interaction down, there was an effect of OoA in the neutral sentences at test ($\chi^2 = 5.52$, $df = 1$, $p = .019$) but not in the meaningful sentences presented in the final exposure session ($\chi^2 = .11$, $df = 1$, $p = .74$); considering the data another way, there was an effect of sentence-type for late-presented words ($\chi^2 = 3.84$, $df = 1$, $p = .0499$) but not early presented words ($\chi^2 = .61$, $df = 1$, $p = .44$) confirming that while total reading times did not increase significantly from exposure to test for early-presented items, they did for late-presented items.

4. General Discussion

This experiment took a novel approach to investigate the acquisition of new written words. Rather than employ an explicit and exhaustive training regime, new words were encountered in more naturalistic circumstances as participants read meaningful sentences containing novel forms. In addition to this incidental exposure, monitoring eye movements provided an index of learning that was unobtrusive and mirrored normal reading circumstances, i.e. silent reading of connected text with no secondary task. Although previous studies have examined eye movement behaviour when readers encounter a novel word in a sentence (e.g., Chaffin et al., 2001; Lowell & Morris, 2014), the acquisition of novel forms over repeated encounters has not been explored in an eye movement experiment. Our data are clear in showing that with each encounter in the exposure phase, reading

times reduced on all eye movement measures. In line with our prediction, this shows that as novel words became more familiar, they engender shorter processing times, consistent with frequency effects seen in many eye movement studies, including those with children in the early stages of learning to read (Blythe et al., 2009; Hyona & Olson, 1995; Huestegge, Radach, Corbic & Huestegge, 2009; Joseph, Nation & Liversedge, 2013).

In addition to this basic familiarity effect, we also asked whether the order in which words were encountered during exposure influenced learning. The test phase of the experiment provided two ways to address this question: via the surprise memory test in which explicit knowledge about the new words was tapped and by monitoring eye movements as participants read the early vs. late words in neutral test sentences. Overall performance was low in the memory task, but nevertheless, there was a clear effect of order with participants showing significantly higher levels of knowledge for the early items. This finding complements other laboratory demonstrations of OoA effects in lexical processing (e.g., Catling et al., 2013; Izura et al., 2011), and extends them to more naturalistic circumstances in which order effects emerged after just 15 exposures distributed across five days, and without explicit instruction. Reading time data from the test sentences also showed an effect of order, with early words receiving shorter reading times than late words. The magnitude of the OoA effect in reading times was large, averaging 55ms in total reading time. The size of this effect is noteworthy, especially when set against the low levels of performance seen in the explicit memory task. This suggests that order of presentation influences processing time early in the course of acquiring a new word: if we index a word being fully known by performance in the explicit memory test, then it is clear that characteristics of that word are already affecting the ease with which it is processed online before the word is well enough known to support performance on the explicit task. This, along with the gradual decline in reading times seen as the exposure phase progressed, is consistent with partial and incremental growth in knowledge occurring as a function of an individual's experience with each word.

There was no hint of an OoA effect during exposure: reading times reduced equivalently for early vs. late items. Yet at test, early items received shorter reading times than late items. This finding is striking for two reasons. First the two sets of items received identical exposure, both in terms of frequency and semantics – exactly the same sentences were used across early vs. late conditions meaning there was complete control over stimulus properties. Second, the test sentences occurred at the end of the experiment, immediately following the final exposure session; test sentences were not marked in any way and so from the participants' point of view, test sentences were not identifiable as such. There was however a crucial difference between the nature of the sentences read during exposure and at test. Throughout exposure, new words were embedded in meaningful sentences, each of which provided information about the new word. At test however, sentence context was neutral and did not provide any information pertaining to the meaning of the new word. Reading times for early words were not significantly slowed in the test sentences, relative to reading time in the final round of exposure. In contrast, reading times for late words increased from exposure to test. If early words enjoy a processing advantage because of general learning principles, as suggested by computational models (e.g., Ellis & Lambon Ralph, 2000; Monaghan & Ellis, 2010), we can speculate that the early words in our experiment gained a higher quality of lexical status than the late words. A consequence of this might be greater context independence (e.g., Perfetti, 2007; Perfetti & Hart, 2002), evidenced here by shorter reading times of these words (but not late learned words) in neutral sentences, as discussed below.

Note that we do not claim that the effects of OoA would always be restricted to neutral sentences. The neutral context is likely to be particularly important here since, due to the relatively small number of exposures, participants were still using context to boost their emerging representation of word meaning during the training sessions. Thus the comparatively difficult task of lexical access when reading the words in neutral sentences at test was sufficient to reveal an advantage for novel words presented early. We think it likely that if our participants had gained sufficiently more exposure

to these words, as they developed more highly specified, precise and context-independent representations of the target words, the OoA effect would be apparent regardless of sentence type.

The extent to which OoA serves as an accurate laboratory analogue of AoA remains an open question. Clearly, the order effects manipulated over a week in our experiment are quite different in scale to the age differences seen in reading development, even when AoA is measured in beginning readers (Nazir et al., 2003). And, the eye movement signature for early vs. late items in our experiment differs from what is seen when AoA is manipulated in familiar words. We saw an effect of order in the relatively late processing measures of total time, whereas Juhasz and Rayner (2006) reported a small ($M = 10\text{ms}$) but reliable effect of AoA in first fixation durations, indicating differences very early in processing. It is likely that effects were observed in the current experiment only in the later measure of total reading times because participants had not yet formed a complete lexical representation of the new words, and thus the relative difficulty they experienced in processing the late-presented items was not constrained to lexical access but rather reflected increased re-reading of the words after reading the sentence context in search of clues as to its meaning (cf. Chaffin et al., 2001; Lowell & Morris, 2014). Potentially with further exposure, an effect of OoA would emerge in first pass processing, as lexical quality increased. Indeed we did see a non-negligible 23ms advantage for early-presented novel words in gaze durations at test suggesting a trend which might reach significance with increased exposure.

An interesting question, beyond the scope of the current study, is to consider whether the dynamics of learning might change if real but rare and unknown words were used in place of novel words. The choice of nonwords as targets was critical for our experiment as it allowed perfect control over stimulus properties, and provided a guarantee that all items were unknown to all participants. It would however be possible to use a similar methodology to systematically track the learning of real words; this would be interesting for at least two reasons. First, we introduced a relatively large number of similar stimuli (i.e., all bisyllabic, six-letter, CVCCVC strings within the same broad semantic

context). It might be that this actually worked against learning due to interference between stimuli. From this perspective, the learning we observed is even more impressive. Nevertheless, an advantage to using real-but-unfamiliar words is that they are more varied in orthographic form, and they can be embedded in natural language sentence contexts that are both richer and more varied than those used in our experiment. This would provide a more realistic experimental situation in which to track learning, albeit it at the cost of tight control over stimulus properties. Second, and related to this, some research that suggests that OoA may be affected by the perceived novelty of the stimulus (Moore & Valentine, 1998) and it will be interesting to systematically explore this in the context of learning from reading in future work.

What can we learn about the acquisition of new written words from our findings? Our experiment provided perfect control over frequency, thus adding to the evidence base that order influences lexical processing independently of frequency. Although semantic factors are implicated in our effect as words were encountered in meaningful sentences, they cannot account for the difference between early vs. late items as they were identical across conditions. Like Monaghan and Ellis (2010) however, we do not wish to argue against a semantic locus to AoA, merely that a semantic difference is not critical to bring about an order effect in learning.

Finally, and moving away from the specifics of AoA, there is evidence that participants are treating the novel words as potential words (see also Lowell & Morris, 2014). In addition to effects of familiarity and order, we also saw a difference in the processing of words in context vs. neutral sentences, consistent with more re-reading in neutral sentences. This suggests that participants were looking to subsequent portions of the text to help serve lexical access, and that the observation that this was especially marked for the late words demonstrates that the ease of lexical access is influenced by exposure history. Despite the low levels of performance on the explicit memory task, the increase in reading times for the late presented items was large, emphasising the utility of a sensitive online measure of processing, especially when used in conjunction with a more explicit offline measure. Taken

together, our online and offline data suggest that while participants may have felt that they learned little about the novel words in terms of what they could explicitly express, their eye movements betrayed a significant processing advantage for early-presented items after minimal exposure. This is consistent with partial word knowledge occurring incrementally and incidentally over time under naturalistic reading conditions, without the need for explicit and repeated instruction.

Acknowledgements

Portions of these data were presented at the 2011 meeting of the European Conference on Eye Movements in Marseilles, France. Many thanks to Barbara Juhasz and Marc Brysbaert for their insightful comments and suggestions. We also thank Eloise Stark and Lucy Payne for research assistance. The manuscript was prepared while Holly Joseph and Kate Nation were supported by the Economic and Social Research Council, and Elizabeth Wonnacott was supported by the British Academy and Economic and Social Research Council.

References

- Anderson, R. C., & Freebody, P. (1981). Vocabulary knowledge. In J. T. Guthrie (Ed.), *Reading Comprehension and teaching: Research reviews*. Newark, DE: International Reading Association, 1981.
- Baayen, R. H. (2008). *Analyzing linguistic data* (Vol. 505). Cambridge, UK: Cambridge University Press.
- 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.
- Blythe, H. I., Liversedge, S. P., Joseph, H. S. S. L., White, S. J., & Rayner, K. (2009). The uptake of visual information during fixations in reading in children and adults. *Vision Research*, *49*(12), 1583-1591.
- Bonin, P., Barry, C., Méot, A., & Chalard, M. (2004). The influence of age of acquisition in word reading and other tasks: A never ending story? *Journal of Memory and Language*, *50*(4), 456-476.
- Bowers, J. S., 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.
- Henderson, L. M., Weighall, A. R., Brown, H., & Gareth Gaskell, M. (2012). Consolidation of vocabulary is associated with sleep in children. *Developmental Science*, *15*(5), 674-687.
- Brybaert, M., & Ghyselinck, M. (2006). The effect of age of acquisition: Partly frequency related, partly frequency independent. *Visual Cognition*, *13*(7-8), 992-1011.
- Brybaert, M., Wijnendaele, I. V., & Deyne, S. D. (2000). Age-of-acquisition effects in semantic processing tasks. *Acta Psychologica*, *104*(2), 215-226.
- Catling, J., Dent, K., Preece, E., & Johnston, R. (2013). Age-of-acquisition effects in novel picture naming: A laboratory analogue. *The Quarterly Journal of Experimental Psychology*, *66*(9), 1756-1763.
- Chaffin, R., Morris, R. K., & Seely, R. E. (2001). Learning new word meanings from context: a study of eye movements. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *27*(1), 225.
- Coltheart, V., Laxon, V.J. & Keating, C. (1988). Effects of word imageability and age of acquisition on children's reading. *British Journal of Psychology*, *79*, 1-12.
- Cortese, M. J., & Khanna, M. M. (2007). Age of acquisition predicts naming and lexical-decision performance above and beyond 22 other predictor variables: An analysis of 2,342 words. *The Quarterly Journal of Experimental Psychology*, *60*(8), 1072-1082.
- Ellis, A. W., & Lambon Ralph, M. A. (2000). Age of acquisition effects in adult lexical processing reflect loss of plasticity in maturing systems: Insights from connectionist networks. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *26*(5), 1103.
- Ghyselinck, M., Lewis, M. B., & Brybaert, M. (2004). Age of acquisition and the cumulative-frequency hypothesis: A review of the literature and a new multi-task investigation. *Acta Psychologica*, *115*(1), 43-67.

Harm, M. W., & Seidenberg, M. S. (1999). Phonology, reading acquisition, and dyslexia: insights from connectionist models. *Psychological Review*, 106(3), 491.

Henderson, L., Weighall, A., Brown, H., & Gaskell, G. (2013). Online Lexical Competition During Spoken Word Recognition and Word Learning in Children and Adults. *Child Development*, 84(5), 1668–1685.

Huestegge, L., Radach, R., Corbic, D., & Huestegge, S. M. (2009). Oculomotor and linguistic determinants of reading development: A longitudinal study. *Vision Research*, 49(24), 2948-2959.

Hyönä, J., & Olson, R. K. (1995). Eye fixation patterns among dyslexic and normal readers: effects of word-length and word-frequency. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 21(6), 1430-1440.

Inhoff, A. W., & Rayner, K. (1986). Parafoveal word processing during eye fixations in reading: Effects of word frequency. *Perception & Psychophysics*, 40(6), 431-439.

Izura, C., Pérez, M. A., Agallou, E., Wright, V. C., Marín, J., Stadthagen-González, H., & Ellis, A. W. (2011). Age/order of acquisition effects and the cumulative learning of foreign words: A word training study. *Journal of Memory and Language*, 64(1), 32-58.

Jaeger, T. F. (2008). Categorical data analysis: Away from ANOVAs (transformation or not) and towards logit mixed models. *Journal of Memory and Language*, 59(4), 434-446.

Johnston, R. A., & Barry, C. (2006). Age of acquisition and lexical processing. *Visual Cognition*, 13(7-8), 789-845.

Joseph, H. S., Nation, K., & Livversedge, S. P. (2013). Using Eye Movements to Investigate Word Frequency Effects in Children's Sentence Reading. *School Psychology Review*, 42(2).

Juhasz, B. J. (2005). Age-of-acquisition effects in word and picture identification. *Psychological Bulletin*, 131(5), 684.

Juhasz, B. J., & Rayner, K. (2006). The role of age of acquisition and word frequency in reading: Evidence from eye fixation durations. *Visual Cognition*, 13(7-8), 846-863.

Landauer, T. K., & Dumais, S. T. (1997). A solution to Plato's problem: The latent semantic analysis theory of acquisition, induction, and representation of knowledge. *Psychological Review*, 104(2), 211.

Landi, N., Perfetti, C. A., Bolger, D. J., Dunlap, S., & Foorman, B. R. (2006). The role of discourse context in developing word form representations: A paradoxical relation between reading and learning. *Journal of Experimental Child Psychology*, 94(2), 114-133.

Lowell, R., & Morris, R. K. (2014). Word length effects on novel words: Evidence from eye movements. *Attention, Perception, & Psychophysics*, 76(1), 179-189.

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.

Monaghan, P., & Ellis, A. W. (2010). Modeling reading development: Cumulative, incremental learning in a computational model of word naming. *Journal of Memory and Language*, 63(4), 506-525.

Monaghan, J., & Ellis, A. W. (2002). What exactly interacts with spelling--sound consistency in word naming? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 28(1), 183.

Moore, V. & Valentine, T. (1998). The effect of age of acquisition on speed and accuracy of naming famous faces. *The Quarterly Journal of Experimental Psychology: Section A*, 51(3), 485-513.

Morrison, C. M., & Ellis, A. W. (1995). Roles of word frequency and age of acquisition in word naming and lexical decision. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 21(1), 116.

McKay, A., Davis, C., Savage, G., & Castles, A. (2008). Semantic involvement in reading aloud: evidence from a nonword training study. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 34(6), 1495.

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.

Nazir, T. A., Decoppet, N., & Aghababian, V. (2003). On the origins of age-of-acquisition effects in the perception of printed words. *Developmental Science*, 6(2), 143-150.

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.

Perfetti, C. (2007). Reading ability: Lexical quality to comprehension. *Scientific Studies of Reading*, 11(4), 357-383.

Perfetti, C. A., & Hart, L. (2002). The lexical quality hypothesis. *Precursors of functional literacy*, 11, 189-214.

Qiao, X., & Forster, K. I. (2012). Novel Word Lexicalization and the Prime Lexicality Effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 39(4), 1064-1074.

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

Quené, H., & Van den Bergh, H. (2008). Examples of mixed-effects modeling with crossed random effects and with binomial data. *Journal of Memory and Language*, 59(4), 413-425.

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

- Rayner, K. (2009). Eye movements and attention in reading, scene perception, and visual search. *The Quarterly Journal of Experimental Psychology*, 62(8), 1457-1506.
- Rayner, K., & Duffy, S. A. (1986). Lexical complexity and fixation times in reading: Effects of word frequency, verb complexity, and lexical ambiguity. *Memory & Cognition*, 14(3), 191-201.
- Ricketts, J., Bishop, D. V., Pimperton, H., & Nation, K. (2011). The role of self-teaching in learning orthographic and semantic aspects of new words. *Scientific Studies of Reading*, 15(1), 47-70.
- Share, D. L. (2008). Orthographic learning, phonological recoding, and self-teaching. *Advances in Child Development and Behavior*, 36, 31-82.
- 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.
- Share, D. L. (1995). Phonological recoding and self-teaching: *sine qua non* of reading acquisition. *Cognition*, 55(2), 151-218.
- Stewart, N., & Ellis, A. W. (2008). Order of acquisition in learning perceptual categories: A laboratory analogue of the age-of-acquisition effect? *Psychonomic Bulletin & Review*, 15(1), 70-74.
- Steyvers, M., & Tenenbaum, J. B. (2005). The Large-Scale Structure of Semantic Networks: Statistical Analyses and a Model of Semantic Growth. *Cognitive Science*, 29(1), 41-78.
- Tamminen, J., & Gaskell, M. G. (2013). Novel word integration in the mental lexicon: Evidence from unmasked and masked semantic priming. *The Quarterly Journal of Experimental Psychology*, 66(5), 1001-1025
- Tamminen, J., & Gaskell, M. G. (2008). Newly learned spoken words show long-term lexical competition effects. *The Quarterly Journal of Experimental Psychology*, 61(3), 361-371.
- 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.
- 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.
- Wang, H. C., Castles, A., & Nickels, L. (2012). Word regularity affects orthographic learning. *The Quarterly Journal of Experimental Psychology*, 65(5), 856-864.
- 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.

Appendix A

Example of experimental sentences in the exposure sessions (Days 1-5) for one of the 16 nonwords (wefdof).

1. It is said that a **wefdof** is a type of hunting pouch
2. Furthermore a **wefdof** is only used by this specific tribe
3. We know that a **wefdof** is approximately the size of a handbag
4. Without anything in it a **wefdof** is quite light
5. Generally a **wefdof** is square in shape
6. Normally a **wefdof** is made from buffalo skin
7. Given the material used to make it a **wefdof** is dark red in colour
8. At the end of a **wefdof** there is a long lip
9. When hunting a **wefdof** will be carried by the men of the tribe
10. Knives, herbs and other hunting materials are stored in a **wefdof** when being used
11. Tribe members place their **wefdof** around their waist
12. The lip on a **wefdof** is used to tuck it into the user's belt
13. Tribe members will wash the dirt from a **wefdof** after they have used it
14. Tribe members will carry a **wefdof** on their person for most of the day
15. Males of the tribe receive a **wefdof** when they come of age and don't replace it

Experimental sentences in the test session (Day 5) for all nonwords

1. The boy ran past the wefdof and then fell over
2. The dog sniffed the hebnin curiously
3. The bird landed on the lisbaw and started singing
4. Nobody had seen the lamcuk for several weeks
5. Under the tree there was a sorlas covered in leaves
6. Someone had broken the leskud but they didn't know who
7. The child stared at the zomlal and thought about what had happened
8. She could see the sabrup lying in the grass
9. All that remained was the gadwob next to him
10. They walked away from the gogruc talking loudly
11. Someone had covered the felror in mud
12. She wasn't sure why the helvof had been left there
13. The old man prodded the musdit with his stick
14. The sight of the broken guvpaf angered the young man
15. Next to the hosfot there was an animal sleeping
16. They had forgotten to take the lurcor with them

Appendix B

Surprise memory test

hunting spear, carving knife, toothbrush, bed, child's toy, tent

dancing equipment, brush, spade, hunting pouch

bucket, raft, fishing net, drinking cup, cooking pot, cart

lamcuk

sorlas

musdit

felror

hebnin

helvof

gvpaf

wefdof

lisbaw

leskud

hosfot

lurcor

zomlal

gadwob

gogruc

sabrup