CAMBRIDGE UNIVERSITY PRESS

The role of oral vocabulary when L2 speakers read novel words: A complex word training study

Ali Behzadnia[1,2,3], Signy Wegener[4], Audrey Bürki[2] and Elisabeth Beyersmann[1]

[1]School of Psychological Sciences, Macquarie University, Sydney, Australia; [2]Department of Linguistics, University of Potsdam, Potsdam, Germany; [3]International Doctorate for Experimental Approaches to Language and Brain (IDEALAB): Universities of Potsdam (DE), Groningen (NL), Newcastle (UK), and Macquarie University (Sydney, AU) and [4]Australian Centre for the Advancement of Literacy, Australian Catholic University, Sydney, Australia

Abstract

The present study asked whether oral vocabulary training can facilitate reading in a second language (L2). Fifty L2 speakers of English received oral training over three days on complex novel words, with predictable and unpredictable spellings, composed of novel stems and existing suffixes (i.e., *vishing*, *vishes*, *vished*). After training, participants read the novel word stems for the first time (i.e., trained and untrained), embedded in sentences, and their eye movements were monitored. The eye-tracking data revealed shorter looking times for trained than untrained stems, and for stems with predictable than unpredictable spellings. In contrast to monolingual speakers of English, the interaction between training and spelling predictability was not significant, suggesting that L2 speakers did not generate orthographic skeletons that were robust enough to affect their eye-movement behaviour when seeing the trained novel words for the first time in print.

Introduction

A positive association between oral vocabulary and reading has been reported in monolingual children and adults (Duff & Hulme, 2012; Hogaboam & Perfetti, 1978; Johnston et al., 2004; McKague et al., 2001), such that prior knowledge of the spoken form of a word conveys a reading accuracy and efficiency advantage over orally unfamiliar words. Recently, it has been proposed that readers might be able to use their knowledge of phoneme-to-grapheme correspondences to form orthographic skeletons, which are expectations of the spellings of words held in oral vocabulary that have not been seen in writing (Beyersmann et al., 2021, 2022a; Jevtović et al., 2022, 2023; Wegener et al., 2018, 2020). The current study builds on prior evidence from monolinguals by adopting a training study design to address the question of whether German–English bilinguals also generate orthographic skeletons when acquiring morphologically complex novel words in their second language (L2).

In a recent review, Wegener et al. (2022a) outlined evidence for the existence of an association between oral vocabulary knowledge and word reading. They noted that the association is supported by results from a range of research designs, including cross-sectional studies of individual differences (e.g., Bowey & Rutherford, 2007; Goff et al., 2005; Ouellette, 2006; Ouellette & Beers, 2010), cross-sectional item-level analyses (Kearns & Al Ghanem, 2019; Nation & Cocksey, 2009; Ricketts et al., 2016), longitudinal studies (Duff et al., 2015; Lee, 2011) and training studies. Of these, training studies provide the strongest evidence for the existence of causal effects (Hulme & Snowling, 2013). Training studies involve two stages: a teaching phase and a testing phase. In the teaching phase, participants are taught either the pronunciation alone or the pronunciations and meanings of new spoken words. In the subsequent testing phase, participants read the trained words and a matched set of untrained words for the first time and their reading accuracy and/or efficiency are recorded. Training studies with both children and adults (Beyersmann et al., 2022a; Duff & Hulme, 2012; Hogaboam & Perfetti, 1978; McKague et al., 2001; Taylor et al., 2011) have found that prior knowledge of the spoken form of a word conveys a reading advantage over untrained words, consistent with the relationship between spoken word knowledge and reading being causal.

A potential cognitive mechanism supporting this link between oral vocabulary and reading was originally proposed, but not tested, by Stuart and Coltheart (1988). These authors suggested that spoken word knowledge might influence word reading before visual exposure to printed word forms, arguing that if a child could segment spoken speech sounds and had some knowledge of letter sounds, they might begin to construct an orthographic lexicon

This article has earned badges for transparent research practices: Open Data and Open Materials. For details see the Data Availability Statement.

prior to the commencement of formal reading instruction. Some years later, Johnston et al. (2004) provided skilled readers with training in novel spoken words before showing the written form for the first time within a masked priming task and found a pattern of results consistent with automatic activation of orthography. Subsequently, and also using the masked priming paradigm, McKague et al. (2008) proposed and found evidence for their CONSONANT FRAME HYPOTHESIS, according to which skilled readers can build under-specified orthographic representations around the consonants of known spoken words that are visually novel. This early work with skilled readers suggested that it was plausible that oral vocabulary knowledge might support word reading before visual exposure.

Drawing on this prior work, Wegener et al. (2018) proposed the ORTHOGRAPHIC SKELETON HYPOTHESIS, according to which, once children have a reasonable appreciation of the mappings between phonemes and graphemes, they are in a position to draw on their oral vocabulary knowledge to form expectations of the spellings of words they have not yet seen in writing. In an initial test of this theory with developing readers, Wegener et al. (2018) taught children in Grade 4 the pronunciations and meanings of novel words in English through oral description of a series of inventions. Next, the trained words, as well as a matched set of untrained words, were embedded in sentences; and children read them for the first time while their eye movements were monitored. Importantly, some novel words had predictable spellings from phonology (e.g., the spoken word "vish" was written as *vish*) and some had unpredictable spellings (e.g., the spoken word "jeab" was written as *jeabb*). The key result was an interaction between training and spelling predictability, such that there was a larger spelling predictability effect for trained compared to untrained novel words. This result was taken to indicate that children generated orthographic skeletons of orally known words, which was evident at the first orthographic exposure. In a follow-up study, Wegener et al. (2020) replicated this finding at the first orthographic exposure and found evidence that children's orthographic skeletons are tentative initial orthographic expectations that are updated as experience with the written word form accrues.

Similarly, skilled readers have been found to generate orthographic skeletons during isolated word reading (Wegener, Wang, et al., 2022b). In an experiment with native Spanish speakers, Jevtović et al. (2022) trained their participants in the pronunciations of a set of novel words. When the written form of the trained and untrained words was presented in a sentence reading task, some items had consistent spellings (only one spelling was possible), while others had inconsistent spellings (two spellings were possible, and participants either saw a preferred or an unpreferred spelling). The results showed that participants read trained consistent and inconsistent words with preferred spellings faster than inconsistent words with unpreferred spellings, which is in line with the hypothesis that Spanish native speakers generated orthographic skeletons during oral word training. Similar findings have recently been reported among French native speakers (Jevtović et al., 2023).

There are two studies that are particularly relevant for the aims of the current research, showing that orthographic skeletons are not limited to mono-morphemic words, but also yield robust facilitation effects for novel stems that are embedded in morphologically complex words during oral training, including English-speaking children (Beyersmann et al., 2022b) and adults (Beyersmann et al., 2021). Using a similar training study design to Wegener et al.'s prior work (Wegener et al., 2018, 2020), adults

were first taught the spoken form of a set of novel morphologically complex words composed of a novel stem and a suffix (e.g., *vish + ing, vish + es, vish + ed*). Next, the adults read the stems of the novel words with predictable and unpredictable spellings from phonology embedded in sentences while their eye movements were monitored. Participants' eye movements revealed the same key interaction between training and spelling predictability as was reported by Wegener et al. (2018, 2020), suggesting that participants formed early orthographic skeletons of the embedded stems which influenced their word reading. These findings indicate that the acquisition of spoken words induces orthographic predictions that go beyond the trained whole words themselves, including setting up orthographic skeletons for stems embedded in morphologically complex words.

What is less clear from this prior research is whether or not similar mechanisms apply during the acquisition of morphologically complex novel words in L2 speakers of English, which was the focus of the current investigation. The formation of orthographic skeletons of novel stems embedded in morphologically complex words must not only involve the process of decomposing the novel words into morphemic subunits (e.g., *vish + ing*), but also the process of generating orthographic skeletons of the embedded novel word stems. While prior research has shown that monolinguals master these complex skills, it is uncertain if L2 speakers have the necessary morphological parsing and phoneme-to-grapheme mapping skills to benefit from spoken language training in similar ways. In particular, morphological processing has been shown to differ across first-language (L1) and second-language speakers. For example, in Silva and Clahsen's study (2008) a masked morphological priming task was used to investigate processing of derived and inflected words in L1 and L2 speakers of English (Chinese, German, or Japanese). They observed priming effects for derived words in both L1 and L2 speakers but priming effects for regular past tense inflections (i.e., *-ed*) were only observed in L1 speakers. In another study, Neubauer and Clahsen (2009) reported the priming facilitation for German regular inflection in both L1 German speakers and advanced L2 speakers of German with Polish as their L1. However, there was an absence of facilitation for irregular inflection in L2 speakers. While several studies reported processing inflectional morphology as challenging for L2 speakers (e.g., Chen et al., 2007; Friederici, 2002; McDonald, 2006; Sabourin & Haverkort, 2003), other studies revealed priming effects for inflections in both L1 and L2 speakers (Heyer & Clahsen, 2015; Jacob et al., 2018; Reifegerste et al., 2019). Based on L2 speakers' particular difficulty in processing inflections, coupled with their potentially less robust phoneme-to-grapheme mapping skills, it is not clear if the previously reported orthographic skeleton effect in monolinguals would generalize to second language speakers of English, a question we aimed to address in the present study.

The present study

We tested whether prior exposure to the oral forms of morphologically complex novel words benefits reading performance in a group of German (L1) – English (L2) bilinguals, building on Beyersmann et al.'s (2021) results from English monolinguals. Testing whether L2 speakers of English can benefit from phoneme-to-grapheme mappings and generate L2 orthographic skeletons of novel words has not been explored previously. The case of German–English bilinguals is particularly intriguing because although German is classified as a shallow orthography

(i.e., highly consistent grapheme-to-phoneme mappings) as opposed to English, which falls at the deep end of the orthographic transparency spectrum, German is less consistent in its phoneme-to-grapheme mappings as there are often several graphemic options to describe the same phoneme. However, despite the inconsistency of the German spelling system, phoneme-grapheme correspondences are typically regular, as opposed to the highly irregular correspondences in the English language. As a result, native speakers of German may rely more heavily on orthographic skeletons during spoken language exposure compared to native speakers of English. In other words, if the current study were to reveal differences between the orthographic skeleton effect in German–English bilinguals compared to the previously observed findings in English monolinguals, it would not likely be due to their lack of general proficiency in predicting spellings from sound in their L1, but more likely attributable to their greater reliance on regularities within the German compared to the English spelling system.

A three-day training study was conducted, following Beyersmann et al.'s experimental design. L2 speakers of English received oral training on a set of inflected novel words over three consecutive days (e.g., *vishing, vished, vishes*). The embedded novel words had predictable or unpredictable spellings from phonology. On the third day, participants took part in an eye-tracking experiment where their eye movements were measured while reading the trained and untrained stems embedded in sentences. We hypothesized that if L2 speakers of English can generate L2 orthographic skeletons of embedded stems during oral word learning and use these in a subsequent reading task, we would expect to replicate the training by spelling-predictability interaction that has been previously reported in native speakers (Beyersmann et al., 2021). That is, we would expect shorter looking times for orally trained items with predictable spellings than unpredictable spellings, and we would expect this difference to be larger than the corresponding difference for untrained items. We expect to observe this interaction at gaze duration, total reading time and regressions in. If the interaction is apparent at gaze duration as we anticipate, then this would likely reflect lexical identification processes, whereas if the interaction is apparent only at total reading time and regressions in, this would suggest that the effect likely reflects lexical integration processes (Rayner & Liversedge, 2011). In contrast, if L2 speakers do not generate L2 orthographic skeletons we would expect no interaction between training and spelling predictability. We pre-registered these hypotheses (https://osf.io/qh7sm).

Method

Participants

First, a power analysis was conducted based on previous data from English monolinguals (Beyersmann et al., 2021). In their study, 40 participants were recruited: they observed training by spelling predictability interaction for the gaze duration, reading aloud and regressions in variables. To estimate the number of participants required to achieve a statistical power of at least 80% we simulated many datasets using the estimates of fixed and random effects from Beyersmann et al. (2021). For each simulated dataset we ran the statistical model that we also applied to analyse our data (see below). For each dependent variable, we looked at the percentage of datasets with significant outcomes for different number of participants. The results of the power analysis showed

more than 80% chance of replicating the interaction effect with 40 participants for the reading aloud and regressions in variables and 70% chance for gaze duration. Accordingly, for the present study fifty L2 English speakers were recruited. All participants were German university students (Mean Age: 23.6, SD: 3.9) from Potsdam University, Germany, who spoke German as their first language. They participated either for course credits or monetary reimbursement. The Language Experience and Proficiency Questionnaire (LEAP-Q; Marian et al., 2007) was used to acquire detailed information about participants' language background (see Table 1). Participants reported high L2 proficiency and

Table 1. Self-Rate Language Experience and Proficiency Questionnaire (LEAP-Q)

L2 measures	L2 self-rates		
	Mean	SD	Range
Age of acquisition	7.83	2.34	1> years
Language ability			
Language preference			
Speaking	26.58%	21.72	0–100%
Reading	29.52%	22.88	0–100%
Language proficiency[a]			
Speaking	7.22	1.11	0–10
Understanding spoken L2	7.80	0.92	0–10
Reading	8.08	1.02	0–10
Contribution to L2 learning[b]			
Friends	5.72	3.28	0–10
Family	2.76	3.40	0–10
Reading	7.90	1.69	0–10
Self-instruction	5.48	2.98	0–10
TV	6.44	3.22	0–10
Radio	4.18	3.55	0–10
Overall L2 exposure	24.8%	17.67	0–100%
Extent of L2 exposure[c]			
Friends	3.92	2.58	0–10
Family	1.32	2.35	0–10
TV	5.84	2.25	0–10
Radio/music	6.68	2.67	0–10
Reading	6.34	2.06	0–10
Self-instruction	2.72	2.63	0–10
L2 Accent[d]	4.07	1.56	0–10
Perceived by self	4.74	1.56	0–10
Immersion to English			
In a country	13.82	38.02	1>months
In a family	29.04	58.75	1>months
In a school/working place	25.68	40.90	1>months

Note: self-reports on L2 (English) history measure
[a]Range: 0 (none) to 10 (perfect)
[b]Range: 0 (not a contributor) to 10 (most important contributor)
[c]Range: 0 (never) to 10 (always)
[d]Range: 0 (none) to 10 (pervasive)

early L2 language acquisition. They reported that reading and exposure to media had the highest rate of contribution to their L2 acquisition. Only some of the participants reported a few months of immersion to an English-speaking environment.

Materials

Novel words

A list of 32 three-phoneme monosyllabic nonwords (e.g., *vish*) were adapted from Wegener et al. (2018) and were identical to those used by Beyersmann et al. (2021). Three morphologically complex novel words were created from each of these nonwords by combining them with three different existing English suffixes (e.g., *vishing*, *vished*, *vishes*). The syntactic function of the novel words remained consistent, meaning that they were always used as verbs during the oral training phase. As is the case for existing English words, the pronunciation of the suffixes, depended on the phonological context (e.g., the past-tense suffix /t/ was added to stems ending in voiceless consonants, and /d/ to stems ending in voiced consonants).

Half of the novel words had a highly predictable spelling from their phonology (e.g., 'b' for /b/ as in *yab*) since they were assigned spellings with frequent phoneme to grapheme mappings. The other half were assigned spellings that were unpredictable as they contained less frequent mappings (e.g., 'bb' for /b/ as in *jeabb*). The frequencies of phoneme to grapheme mappings were extracted from the CELEX database (Baayen et al., 1993). Bigram and trigram frequencies were extracted from the English Lexicon Project (ELP; Balota et al., 2007) and SUBTLEX-DE databases (Brysbaert et al., 2011). The predictable trained and predictable untrained items, as well as the unpredictable trained and unpredictable untrained items, were matched on number of letters, English and German number of phonemes, English and German number of syllables, English and German logarithmic bigram frequency, and English and German logarithmic trigram frequency (see Table 2). Although the words were different in terms of their spelling predictability, they could be read aloud correctly using the most common grapheme-phoneme mappings.

Two sets of 16 novel words were used for oral training. All the novel words started with a consonant followed by a vowel. Half of the participants were trained with one set; the other half were trained with the other set. Both sets had novel words with predictable and unpredictable spellings. See Appendix A (Beyersmann et al., 2021, p. 97) for the full lists of novel words.

Eye-tracking sentences

Thirty-two sentences were created, one sentence for each of the novel words. There were also eight filler sentences which contained additional untrained word stems. In all sentences, only the stem was used (e.g., *vish*) and was embedded in the middle of the sentence (see Appendix B; Beyersmann et al., 2021, p. 98). The target words were placed in a position such that they were predictable for meaning (see Beyersmann et al., 2021; Wegener et al., 2018). For example: *Ben put the machine into the fish tank to* <u>chig</u> *the glass clean again.*

Procedure

Oral novel word training and testing took place over three consecutive days. Each training session on Day 1 and 2 lasted about 30 minutes. Day 3 took about 90 minutes as the participants completed the training phase and performed post training

Table 2. English and German Item Characteristics across Novel Word Sets and Spelling Predictability Conditions

Measures	Set 1		Set 2	
	Predictable			
	Mean	SD	Mean	SD
Number of letters	3.5	0.53	3.5	0.53
English number of phonemes	3	0	3	0
English number of syllables	1	0	1	0
English logarithmic bigram frequency	3.49	0.39	3.72	0.22
English logarithmic trigram frequency	1.86	0.82	2.00	0.89
German number of phonemes	3	0	3	0
German number of syllables	1	0	1	0
German logarithmic bigram frequency	3.83	0.52	4.17	0.38
German logarithmic trigram frequency	1.59	0.97	2.11	0.93
	Unpredictable			
Number of letters	4.5	0.53	4.5	0.53
English number of phonemes	3	0	3	0
English number of syllables	1	0	1	0
English logarithmic bigram frequency	3.52	0.39	3.55	0.33
English logarithmic trigram frequency	1.88	0.75	1.78	0.71
German number of phonemes	3.63	0.52	3.63	0.52
German number of syllables	1.50	0.53	1.50	0.53
German logarithmic bigram frequency	3.93	0.33	3.88	0.47
German logarithmic trigram frequency	2.10	1.07	1.79	0.78

tasks which are summarized below (see Table 3; Beyersmann et al., 2021, p. 90). Throughout the training and post-training sessions, participants were instructed that they were presented with English novel words and that their task was to read aloud the English novel words. All experimental instructions were provided in English, and throughout the entire testing session the experimenter and participants communicated in the English language only.

Training phase

The oral training phase closely followed the procedure by Beyersmann et al. (2021). Participants received oral training of novel words in small groups of 2–4 students. They were informed that they would be learning novel English words about 'Professor Parsnip's Inventions' and learn about the features and functions of the inventions. The participants were shown pictures of several inventions and received oral descriptions of each invention using the novel words. The written form of the trained words was never shown to the participants. For instance, they learned that

Table 3. Testing Procedure Involving Three Consecutive Days of Oral Vocabulary Training

Test components		Materials	Time
Day 1	LEAP Questionnaire Oral Vocabulary training Rehearsal 1 Rehearsal 2	8 words	30 min.
Day 2	Oral Vocabulary training Rehearsal 1 Rehearsal 2	8 words	30 min.
Day 3	Oral Vocabulary training Rehearsal Picture-naming test Eye tracking experiment Reading aloud Spelling task	all 16 words	90 min

Note. From "Learning morphologically complex spoken words: Orthographic expectations of embedded stems are formed prior to print exposure," by Beyersmann et al., 2021, *Journal of Experimental Psychology: Learning, Memory, and Cognition, 47(1)*, p. 90 (https://doi.org/10.1037/xlm0000808)

"Professor Parsnip has invented a machine that *chigs*. It is used for cleaning out fish tanks. It has a sponge and is shaped like an arm"; see Figure 1).

On Day 1, the participants were orally trained twice on eight complex novel words with half having a predictable and half an unpredictable spelling. The remaining eight words were introduced and also trained twice on Day 2. On Day 3, all 16 trained novel words and the corresponding features and functions of the inventions were reviewed.

Post-training phase: Picture naming.
This task was performed to check whether the participants learned the novel words and their meanings. Therefore, the participants were shown the picture of each of the inventions and asked about their function (e.g., cleans fish tanks) and usage (e.g., it *chigs*). Participants' responses were recorded.

Post-training phase: Eye-tracking experiment.
The participants had the first orthographic exposure to the novel words stems during sentence reading. Sentences were presented
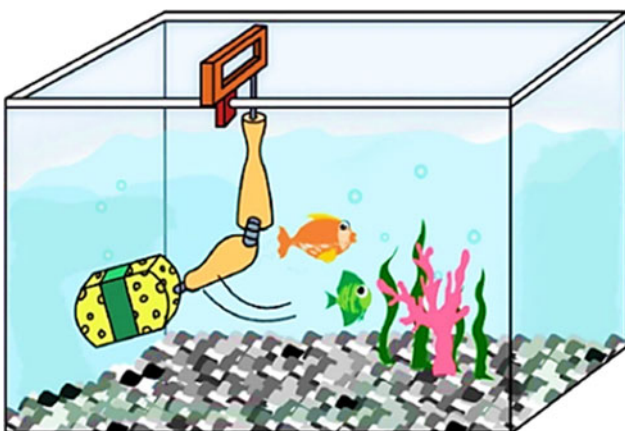

Figure 1. Example of a Picture used During Oral Novel Word Training A machine that is used to 'chig' fish tanks.

one by one on the computer screen. Each sentence appeared on a single line. Eye movements were monitored while participants read the sentences silently. The participants read sentences on a computer monitor at a viewing distance of 85 cm. Each character covered 0.26 of horizontal visual angle. Sentences were presented in black, Courier New font on a white background. Eye movements were recorded using an EyeLink 1000 desk-mounted eye tracker (SR Research; Mississauga, Canada) with a sampling rate of 1000 Hz. Participants' right-eye movements were recorded during binocular reading.

A nine-point calibration procedure was performed. Participants fixated a drift correct target prior to each trial. The experiment began with three practice sentences. For all trials, participants ended each trial by fixating box on the right, underneath each sentence. To promote attention to task, participants were asked a yes/no question after each trial.

Participants' reading of the target words was captured by four eye movement dependent variables: first fixation duration (duration of initial fixation on the target word); gaze duration (sum of all fixations made on the target before the eyes move past the target word to a subsequent word within the sentence); total reading time (sum of all fixations on the target word, including any regressions back to it); and regressions in (probability of making a regression back to the target word from a later portion in the sentence).

Post-training phase: Reading aloud.
Participants read aloud all trained (n = 16) and untrained (n = 16) word stems presented individually in a randomized order in the center of a computer screen using DMDX software (Forster & Forster, 2003). Each trial consisted of an 800-ms fixation cross followed by the target word which remained on the screen until 2 seconds had elapsed. Participants were instructed to read aloud each word as quickly and accurately as possible.

Post-training phase: Spelling.
The experimenter read aloud the trained and untrained novel word stems and the participants were required to spell the novel words as they were written in the eye-tracking and the reading aloud tasks.

Results

Analysis
We investigated the effect of training and spelling predictability on four eye-tracking dependent variables, including first fixation duration, gaze duration, total reading time, and regressions in. For continuous variables, extremely short or long (below 80 milliseconds and above 1200 milliseconds) looking times were removed. The distribution of these variables was visualized using R statistical software (R Core Team, 2020). Outliers were detected using a density plot. Extreme influential observations were identified using the influence.ME package (Nieuwenhuis et al., 2012) and removed (i.e., 40 out of 1462 trials for first fixation duration, 41 trials for gaze duration, and 19 trials for total reading time). We report the results of the models without influential observations. Following Box-Cox tests (Box & Cox, 1964), the continuous variables were log transformed. The lme4 package (Bates et al., 2015) was used to run the statistical models.

One model was run for each dependent variable: first fixation duration, gaze duration, total reading time, and regressions in. All models had training (trained condition was coded as −0.5 and

untrained condition was coded as 0.5), spelling predictability (predictable condition was coded as −0.5 and unpredictable condition was coded as 0.5), and their interaction as fixed-effects. The random-effects structure included by-participant and by-item varying intercepts and slopes. The initial model had no correlation between intercepts and slopes. When convergence issues occurred, random slopes were removed one by one, starting with the random slope with the smallest value. To analyze regressions in, we ran a generalized linear mixed-effects model with the glmer function in the lme4 R package. Outliers and extreme influential observations were removed from the dataset (i.e., 58 out of 2453 trials for first fixation duration, 70 trials for gaze duration, and 60 trials for total reading time).

Similarly, we investigated the effect of training and spelling predictability on participants' reading aloud (i.e., response times and error rates) and spelling responses (i.e., error rates). The statistical models were the same as in the eye-tracking analyses. To analyze the picture naming task, a generalized linear mixed-effects model was run with response accuracy as the dependent variable. The model had word set (Set 1 was coded as −0.5 and Set 2 was coded as 0.5), spelling predictability (predictable condition was coded as −0.5 and unpredictable condition was coded as 0.5), and their interaction as fixed-effects. The same statistical models were used as in the analyses of the previous tasks.

Picture naming task

Participants were highly accurate (i.e., 97.5%, $SD = 0.15$) in recalling the orally trained novel words. The statistical model showed no significant effect of word set ($\beta = -0.22$, $SE = 1.86$, $z = -0.12$, $p = 0.904$), spelling predictability ($\beta = -0.82$, $SE = 0.83$, $z = -1$,

$p = 0.321$), or interaction between word set and spelling predictability ($\beta = 0.95$, $SE = 1.62$, $z = 0.58$, $p = 0.559$). The results show that participants were successful in learning the complex novel words through oral training and were able to correctly associate the novel words with their corresponding pictures of the inventions.

Eye movements

First fixation duration

The statistical model showed no significant effect of training ($\beta = -8.78 \times 10^{-5}$, $SE = 1.72 \times 10^{-2}$, $t = -0.005$, $p = 0.996$), spelling predictability ($\beta = 7.52 \times 10^{-3}$, $SE = 2.37 \times 10^{-2}$, $t = 0.317$, $p = 0.754$), or interaction between training and spelling predictability ($\beta = 3.09 \times 10^{-2}$, $SE = 3.44 \times 10^{-2}$, $t = 0.898$, $p = 0.369$; see Figure 2).

Gaze duration

The statistical model showed no significant effect of training ($\beta = 0.05$, $SE = 0.025$, $t = 1.944$, $p = 0.064$). However, the effect of spelling predictability was significant ($\beta = 0.21$, $SE = 0.032$, $t = 6.448$, $p < .001$). The interaction between training and spelling predictability was not significant ($\beta = -0.01$, $SE = 0.05$, $t = -0.339$, $p = 0.737$), providing no support for or against the hypothesis that spelling predictability influenced looking times differently for trained and untrained target words.

Total reading time

The statistical model revealed significant effects of training ($\beta = 0.24$, $SE = 0.04$, $t = 5.76$, $p < .001$) and spelling predictability ($\beta = 0.31$, $SE = 0.06$, $t = 5.07$, $p < .001$). The results indicate that looking times were shorter for trained than untrained target
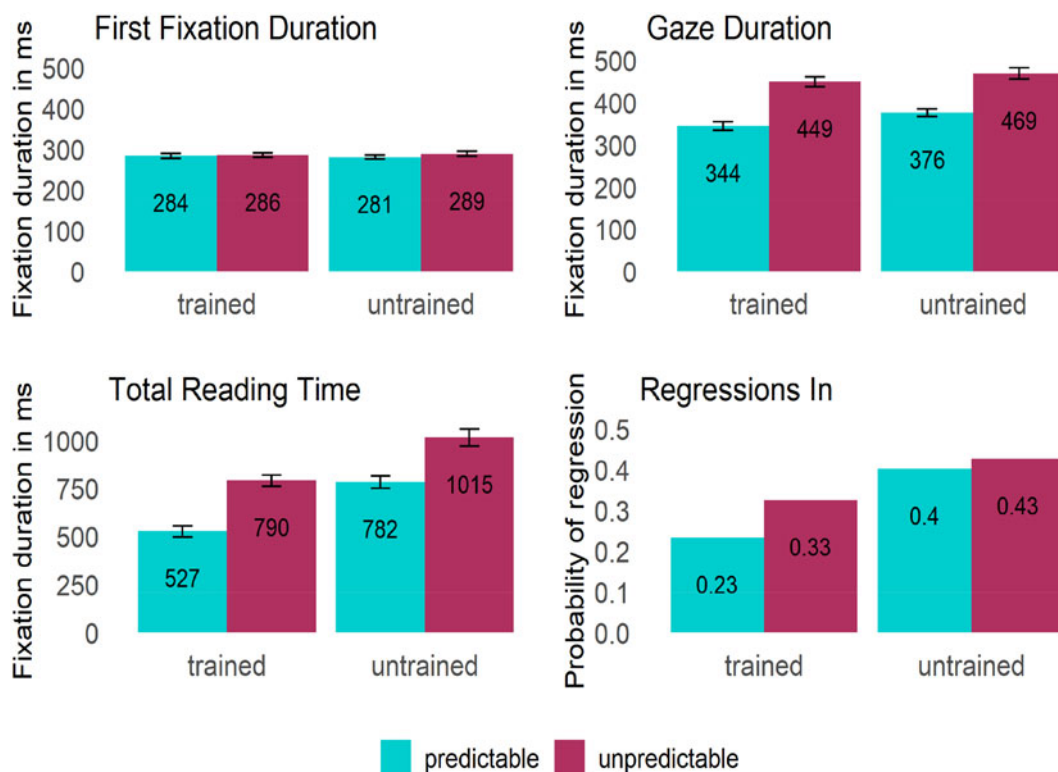

Figure 2. Eye Movements Data of L2 Speakers of English
Note. Means (in milliseconds) and standard errors for first fixation duration, gaze duration, and total reading time are shown in three panels. The probability of regression reflects the likelihood of the occurrence of regression back to the target word.

words and for target words with predictable than unpredictable spellings. The interaction between training and spelling predictability was not significant ($\beta = -0.09$, $SE = 0.085$, $t = -1.066$, $p = 0.295$).

Regressions in

The statistical model revealed a significant effect of training ($\beta = 0.74$, $SE = 0.13$, $z = 5.55$, $p < .001$). The results indicate that probability of regressing to the target word was less for trained than untrained target words. The effect of spelling predictability ($\beta = 0.33$, $SE = 0.177$, $z = 1.868$ $p = 0.061$) and the interaction between training and spelling predictability were not significant ($\beta = -0.40$, $SE = 0.258$, $z = -1.549$, $p = 0.121$).

Reading aloud

Response times

One participant's data were excluded due to technical recording issues. In addition, four participants with error rates above 40% were excluded. Incorrect responses were removed from the dataset (22.3% of all data). Following the Box-Cox test (Box & Cox, 1964), we used the inverse transformation of response times as the dependent variable. The statistical model revealed significant effects of training ($\beta = 6.39 \times 10^{-5}$, $SE = 1.20 \times 10^{-5}$, $t = 5.30$, $p < .001$), spelling predictability ($\beta = 1.25 \times 10^{-4}$, $SE = 3.40 \times 10^{-5}$, $t = 3.69$, $p < .001$), and an interaction between training and spelling predictability ($\beta = 5.75 \times 10^{-5}$, $SE = 2.18 \times 10^{-5}$, $t = 2.63$, $p = 0.01$). This indicates that participants were faster in reading aloud trained items with predictable than unpredictable spellings (see Figure 3).

Error rates

The results revealed a significant effect of training ($\beta = -0.81$, $SE = 0.15$, $z = -5.42$, $p < .001$), indicating that participants made

fewer errors reading trained than untrained items. The effect of spelling predictability ($\beta = -0.47$, $SE = 0.41$, $z = -1.15$, $p = 0.249$) and its interaction with training were not significant ($\beta = -0.28$, $SE = 0.34$, $z = -0.83$, $p = 0.403$).

Spelling

In the spelling data, there was a significant effect of training ($\beta = 0.28$, $SE = 0.117$, $z = 2.44$, $p = 0.014$) showing that participants were more accurate in producing the written form of the trained than untrained novel word stems. In addition, there was a significant effect of spelling predictability ($\beta = 1.47$, $SE = 0.30$, $z = 4.97$, $p < .001$) showing that they were more accurate in producing the written form of the novel word stems with predictable than unpredictable spelling. The interaction between training and spelling predictability was not significant ($\beta = 0.09$, $SE = 0.235$, $z = 0.40$, $p = 0.685$; see Figure 4).

Discussion

The present study used an oral novel word learning paradigm to investigate if L2 speakers of English generate L2 orthographic skeletons during oral training and benefit from them when reading the novel words for the first time. During three consecutive days of training, participants received oral descriptions of different inventions using novel word stems (e.g., *vish*) combined with three different inflectional morphemes (i.e., *vishing*, *vishes*, *vished*). Following training, on the third day, participants read the novel word stems for the first time embedded in sentences while their eye movements were tracked. In addition, they performed a picture naming, a reading aloud, and a spelling task.

The results of the eye-tracking experiment showed a significant effect of training on total reading time and regressions in. This is
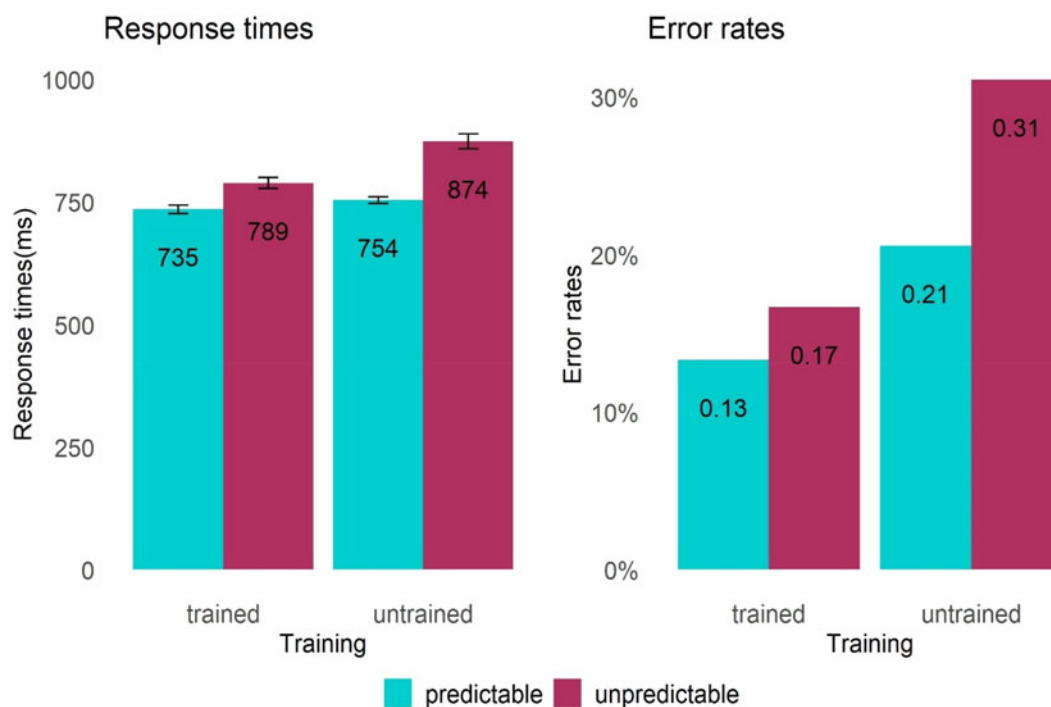

Figure 3. L2 Reading Aloud Data
Note. Mean response times (ms; left panel) and error rates (right panel) for the reading aloud task for the L2 data are presented.

Spelling error rates

Figure 4. L2 Spelling Data
Note. Mean error rates for each condition in the spelling task.

the first key finding showing that adult L2 speakers of English benefited from prior oral vocabulary training when reading novel words for the first time. This is consistent with the main effect of training previously evidenced in English monolinguals (see Beyersmann et al., 2021), although the effect occurs later in the eye movement record of the L2 speakers, perhaps implying somewhat reduced reading efficiency in the L2 sample.

In addition, a significant effect of spelling predictability was observed on gaze duration and total reading time showing that participants spent less time fixating words with predictable than unpredictable spellings. However, although L2 speakers were influenced by spelling predictability, this effect was not modulated by whether the words were trained or untrained. The observed absence of a training by spelling predictability interaction clearly contrasts with the earlier findings from English monolinguals (Beyersmann et al., 2021; Wegener et al., 2018, 2020). To statistically compare the current results from L2 with Beyersmann et al.'s monolingual data, we carried out an additional, pre-registered set of combined analyses across both data sets. The full results are reported in the study's OSF repository: https://osf.io/u7wkq/?view_only = 3d4c8b8230154a2facfe7fbd883d9e23 The combined analyses revealed a significant three-way interaction between training, spelling predictability and the participant group (monolinguals vs. bilinguals) for gaze duration and total reading time. This suggests that although L2 speakers were sensitive to differences between common and unusual spellings in English, as evidenced by a main effect of predictability of the novel words, they did not demonstrate the orthographic skeleton effect observed in monolinguals.

There are several different possibilities why this might have been the case. One possibility is that participants simply did not generate orthographic skeletons at all. While the data do not allow us to rule out this possibility, we consider it unlikely. A second possibility is that participants generated English L2 orthographic skeletons of the novel words during oral learning, but then failed to integrate them when reading the words in the eye-tracking task. Exactly why this might have been the case is not clear, although it could be due to their generally lower levels of reading experience in English compared to native language speakers (Weber & Broersma, 2012). Some studies have reported the impact of L2 proficiency on morphological processing. For instance, Feldman et al. (2010) reported priming effects for inflected verbs in L2 speakers was modulated by the English language proficiency of the participants. That is, the higher the language proficiency was, the stronger the priming effects were.

Likewise, Coughlin and Tremblay (2015) found that priming effects for French inflected words were increased by participants' L2 proficiency. This might explain why L2 speakers failed to generate orthographic skeletons through oral training.

A third possibility is that participants did generate orthographic skeletons, but they were not precisely based on L2 phoneme-grapheme correspondences. We speculate that although participants knew they were learning novel English words, they may have had a greater degree of automaticity in linking phonemes to graphemes in their L1 (German), which may in turn have interfered with the process of generating English L2 orthographic skeletons. For example, in English the phoneme /j/ is spelled as 'y' as in 'yes'; however, in German the same phoneme has a different written form, that is /j/ as in 'ja' (means 'yes' in German). Several other such examples exist too (for instance, the phoneme /f/ in English is written as either 'f' or 'ph' as in 'four'; however, in German the same phoneme can also be written as 'v' like 'vier' (i.e., /fɪɐ/ means 'four' in German). The difference between the German and English orthographies may have had an impact on the creation of orthographic skeletons as well as reading. That is, in English a grapheme corresponds to multiple phonemes (i.e., deep orthography), whereas in German it is more common to have one-to-one correspondence between graphemes and phonemes (i.e., shallow orthography; Liu & Cao, 2016). Therefore, this difference makes it more difficult and time consuming for L2 speakers of English to make accurate and fast correspondences between phonemes and graphemes, which in turn may impact generating orthographic skeletons. Items containing phoneme-grapheme mappings that conflict in German and English were present in the current experiment, across both spelling predictability conditions. If participants automatically drew on their L1 phoneme-to-grapheme mappings, this could have resulted in the formation of an orthographic skeleton that differed from the English phoneme-to-grapheme mappings (e.g., upon hearing the spoken word 'yab' a German L1 speaker may automatically generate the spelling *jab*). If this occurred for items with predictable English spellings, then at least some of the orthographic skeletons participants formed would not have matched the orthographic form they saw, thus potentially limiting the effect of spelling predictability for orally trained items. Some tentative support for this interpretation comes from participants' reading aloud and spelling responses: German grapheme-phoneme correspondences were used when encountering the stems for a proportion of items (for instance, reading aloud 'jit' as /jit/ or spelling 'meaph' as 'meave').

In the reading aloud task, we observed a significant effect of training, spelling predictability and their interaction on reading latencies, and an effect of training on reading aloud accuracy. This task was performed after the eye-tracking task, which means that participants had already encountered the spelling of each word. The results of the reading aloud task indicate that participants benefitted from their prior oral vocabulary knowledge of the trained items, regardless of the predictability of their spellings. This suggests that L2 speakers were able to draw on their oral vocabulary knowledge to assist with the process of forming connections between the orthographic form that participants first encountered during the eye-tracking task and the phonological form that participants had acquired during oral training.

Prior work with both children and adults (Duff & Hulme, 2012; Hogaboam & Perfetti, 1978; Johnston et al., 2004; McKague et al., 2001) has shown that monolingual English

speakers demonstrate a reading accuracy and efficiency advantage for orally trained words compared to untrained words. When a novel word is encountered in print, the reader should phonologically decode the word and, if the word is orally familiar, the reader likely then attempts to match the decoded pronunciation with an entry stored in oral vocabulary. Current theories suggest that there may be two time points at which oral vocabulary knowledge might assist with the process of reading novel words (see Wegener et al., 2022a). The first occurs prior to visual exposure via the generation of orthographic skeletons, as described earlier, which facilitate this matching process. The second occurs from the point of visual exposure via a process termed set for variability (Venezky, 1999) or mispronunciation correction (Dyson et al., 2017), in which decoding attempts undergo some adjustment in order to match them with known spoken words. Given that the current study did not find evidence that L2 learners of English had generated orthographic skeletons prior to visual exposure but they still demonstrated facilitated reading of trained compared to untrained words, this suggests that the benefit of having a word in oral vocabulary may have been conferred via this second process of adjusting decoding attempts. The finding that, on the eye movement measures, the training effect only emerged late in processing supports this interpretation (see Murray et al., 2022).

Further research might investigate whether or not children as beginner L2 learners of English (i.e., German L1 speakers) show similar patterns of learning and are sensitive to the orthographic skeleton of oral words like native children. Moreover, to further explore the challenges of L2 vocabulary acquisition, an interesting extension of the current work would be the direct comparison between novel words that provide an exact match in L1 and L2 phoneme-grapheme correspondences and novel words differing in their L1 and L2 phoneme-grapheme correspondences. It is also worth investigating whether combining novel word stems with derivational morphemes (e.g., *vishist, vishment, vishity*), as opposed to the inflectional forms that were presently used (e.g., *vishing, vished, vishes*) makes a difference in learning and decomposing novel words as well as generating the orthographic skeleton in both L1 and L2 speakers of English. This can contribute to the mixed literature showing differences of morphological priming effects in processing inflected words in L1 vs. L2 but not necessarily in processing derived words (Jacob, 2018; Reifegerste et al., 2019; Silva & Clahsen, 2008). We suspect that German native speakers are naturally more reliant on German than English phoneme-grapheme correspondences, suggesting that the previously evidenced spelling-by-predictability interaction in English monolinguals (Beyersmann et al., 2021; Wegener et al., 2018), Spanish native speakers (Jevtović et al., 2022) and French native speakers (Jevtović et al., 2023) should be replicable in German monolinguals.

In sum, the current findings show that L2 speakers successfully learned the spoken novel words, which led to overall shorter looking times when reading the novel stems embedded in sentences for the first time. Participants also showed sensitivity to the spelling predictability of the novel words, which was however not modulated by oral vocabulary training. Crucially, the eye-movement data revealed that English L2 speakers, as opposed to English monolinguals, did not generate orthographic skeletons that were robust enough to affect their eye-movement behavior when seeing the novel words for the first time in print. Future work may further explore the mechanisms that L2 speakers use in predicting orthographic form based on oral exposure.

References

Baayen, R. H., Piepenbrock, R., & Van Rijn, H. (1993). *The CELEX lexical database (CD-ROM)*.

Balota, D. A., Yap, M. J., Hutchison, K. A., Cortese, M. J., Kessler, B., Loftis, B., Neely, J. H., Nelson, D. L., Simpson, G. B., & Treiman, R. (2007). The English lexicon project. *Behavior Research Methods*, 39(3), 445–459. https://doi.org/10.3758/bf03193014

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67(1), 1–48. https://doi.org/10.18637/jss.v067.i01

Beyersmann, E., Wegener, S., Nation, K., Prokupzcuk, A., Wang, H. C., & Castles, A. (2021). Learning morphologically complex spoken words: Orthographic expectations of embedded stems are formed prior to print exposure. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 47(1), 87–98. https://doi.org/10.1037/xlm0000808

Beyersmann, E., Wegener, S., Pescuma, V. N., Nation, K., Colenbrander, D., & Castles, A. (2022a). The effect of oral vocabulary training on reading novel complex words. *Quarterly Journal of Experimental Psychology*, 0(0). https://doi.org/10.1177/17470218221113949

Beyersmann, E., Wegener, S., Spencer, J., & Castles, A. (2022b). Acquisition of orthographic forms via spoken complex word training. *Psychonomic Bulletin & Review*, 1–12. https://doi.org/10.3758/s13423-022-02185-y

Bowey, J. A., & Rutherford, J. (2007). Imbalanced word-reading profiles in eighth-graders. *Journal of Experimental Child Psychology*, 96(3), 169–196. https://doi.org/10.1016/j.jecp.2006.11.001

Box, G. E. P., & Cox, D. R. (1964). An analysis of transformations. *Journal of the Royal Statistical Society: Series B (Methodological)*, 26(2), 211–243. https://doi.org/10.1111/j.2517-6161.1964.tb00553.x

Brysbaert, M., Buchmeier, M., Conrad, M., Jacobs, A. M., Bölte, J., & Böhl, A. (2011). The word frequency effect: A Review of Recent Developments and Implications for the Choice of Frequency Estimates in German. *Experimental Psychology*, 58(5), 412–424. https://doi.org/10.1027/1618-3169/a000123

Chen, L., Shu, H. U. A., Liu, Y., Zhao, J., & Li, P. (2007). ERP signatures of subject–verb agreement in L2 learning. *Bilingualism: Language and Cognition*, 10(2), 161–174. https://doi.org/10.1017/S136672890700291X

Coughlin, C. E., & Tremblay, A. (2015). Morphological decomposition in native and non-native French speakers. *Bilingualism: Language and Cognition*, 18(3), 524–542. https://doi.org/10.1017/S1366728914000200

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. https://doi.org/10.1080/10888438.2011.598199

Duff, F. J., Reen, G., Plunkett, K., & Nation, K. (2015). Do infant vocabulary skills predict school-age language and literacy outcomes? *Journal of Child Psychology and Psychiatry*, 56(8), 848–856. https://doi.org/10.1111/jcpp.12378

Dyson, H., Best, W., Solity, J., & Hulme, C. (2017). Training mispronunciation correction and word meanings improves children's ability to learn to read words. *Scientific Studies of Reading*, 21(5), 392-407. https://doi.org/10.1080/10888438.2017.1315424

Feldman, L. B., KostiĆ, A., Basnight-Brown, D. M., ĐUrĐEviĆ, D. F., & Pastizzo, M. J. (2010). Morphological facilitation for regular and irregular verb formations in native and non-native speakers: Little evidence for two distinct mechanisms. *Bilingualism: Language and Cognition*, 13(2), 119–135. https://doi.org/10.1017/S1366728909990459

Forster, K. I., & Forster, J. C. (2003). DMDX: A Windows display program with millisecond accuracy. *Behavior Research Methods, Instruments, & Computers*, 35(1), 116–124. https://doi.org/10.3758/BF03195503

Friederici, A. D. (2002). Towards a neural basis of auditory sentence processing. *Trends in Cognitive Sciences*, 6(2), 78–84. https://doi.org/10.1016/S1364-6613(00)01839-8

Goff, D. A., Pratt, C., & Ong, B. (2005). The relations between children's reading comprehension, working memory, language skills and components of reading decoding in a normal sample. *Reading and Writing*, 18(7), 583–616. https://doi.org/10.1007/s11145-004-7109-0

Heyer, V., & Clahsen, H. (2015). Late bilinguals see a scan in scannerAND in scandal: dissecting formal overlap from morphological priming in the processing of derived words. *Bilingualism: Language and Cognition*, 18(3), 543–550. https://doi.org/10.1017/S1366728914000662

Hogaboam, T. W., & Perfetti, C. A. (1978). Reading skill and the role of verbal experience in decoding. *Journal of Educational Psychology*, 70(5), 717–729. https://doi.org/10.1037/0022-0663.70.5.717

Hulme, C., & Snowling, M. J. (2013). Learning to read: What we know and what we need to understand better. *Child Development Perspectives*, 7(1), 1–5. https://doi.org/10.1111/cdep.12005

Jacob, G. (2018). Morphological priming in bilingualism research. *Bilingualism: Language and Cognition*, 21(3), 443–447. https://doi.org/10.1017/S1366728917000451

Jacob, G., Heyer, V., & Veríssimo, J. (2018). Aiming at the same target: A masked priming study directly comparing derivation and inflection in the second language. *International Journal of Bilingualism*, 22(6), 619–637. https://doi.org/10.1177/1367006916688333

Jevtović, M., Antzaka, A., & Martin, C. D. (2022). Gepo with a G, or Jepo with a J? Skilled readers generate orthographic expectations for novel spoken words even when spelling is uncertain. *Cognitive Science*, 46(3), e13118. https://doi.org/10.1111/cogs.13118

Jevtović, M., Antzaka, A., & Martin, C. D. (2023). Déjà-lu: When orthographic representations are generated in the absence of orthography. *Journal of Cognition*, 6(1), 7. https://doi.org/10.5334/joc.250

Johnston, M., McKague, M., & Pratt, C. (2004). Evidence for an automatic orthographic code in the processing of visually novel word forms. *Language and Cognitive Processes*, 19(2), 273–317. https://doi.org/10.1080/01690960344000189

Kearns, D. M., & Al Ghanem, R. (2019). The role of semantic information in children's word reading: Does meaning affect readers' ability to say polysyllabic words aloud? *Journal of Educational Psychology*, 111, 933–956. https://doi.org/10.1037/edu0000316

Lee, J. (2011). Size matters: Early vocabulary as a predictor of language and literacy competence. *Applied Psycholinguistics*, 32(1), 69–92. https://doi.org/10.1017/s0142716410000299

Liu, H., & Cao, F. (2016). L1 and L2 processing in the bilingual brain: A meta-analysis of neuroimaging studies. *Brain and Language*, 159, 60–73. https://doi.org/10.1016/j.bandl.2016.05.013

Marian, V., Blumenfeld, H. K., & Kaushanskaya, M. (2007). The Language Experience and Proficiency Questionnaire (LEAP-Q): Assessing Language Profiles in Bilinguals and Multilinguals. *Journal of Speech, Language, and Hearing Research*, 50(4), 940–967. https://doi.org/10.1044/1092-4388(2007/067)

McDonald, J. L. (2006). Beyond the critical period: Processing-based explanations for poor grammaticality judgment performance by late second language learners. *Journal of Memory and Language*, 55(3), 381–401. https://doi.org/10.1016/j.jml.2006.06.006

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. https://doi.org/10.1016/S0010-0277(00)00150-5

McKague, M., Davis, C., Pratt, C., & Johnston, M. B. (2008). The role of feedback from phonology to orthography in orthographic learning: an extension of item-based accounts. *Journal of Research in Reading*, 31(1), 55–76. https://doi.org/10.1111/j.1467-9817.2007.00361.x

Murray, L., Wegener, S., Wang, H.-C., Parrila, R., & Castles, A. (2022). Children processing novel irregular and regular words during reading: An eye tracking study. *Scientific Studies of Reading*, 26(5), 417–431. https://doi.org/10.1080/10888438.2022.2030744

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. https://doi.org/10.1016/j.jecp.2009.03.004

Neubauer, K., & Clahsen, H. (2009). Decomposition of inflected words in a second language: An experimental study of german participles. *Studies in Second Language Acquisition*, 31(3), 403–435. https://doi.org/10.1017/S0272263109090354

Nieuwenhuis, R., te Grotenhuis, M., & Pelzer, B. (2012). influence. ME: Tools for Detecting Influential Data in Mixed Effects Models. *R Journal*, 4(2), 38–47. https://doi.org/10.32614/RJ-2012-011

Ouellette, G. (2006). What's meaning got to do with it: The role of vocabulary in word reading and reading comprehension. *Journal of Educational Psychology*, 98, 554–566. https://doi.org/10.1037/0022-0663.98.3.554

Ouellette, G., & Beers, A. (2010). A not-so-simple view of reading: how oral vocabulary and visual-word recognition complicate the story. *Reading and Writing*, 23(2), 189-208. https://doi.org/10.1007/s11145-008-9159-1

Rayner, K., & Liversedge, S. P. (2011). Linguistic and cognitive influences on eye movements during reading. In S. P. Liversedge, I. Gilchrist, & S. Everling (Eds.), *The Oxford Handbook of Eye Movements* (pp. 752–766). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199539789.013.0041

R Core Team. (2020). *R: A language and environment for statistical computing. R Foundation for Statistical Computing*. In Vienna, Austria. https://www.R-project.org/

Reifegerste, J., Elin, K., & Clahsen, H. (2019). Persistent differences between native speakers and late bilinguals: Evidence from inflectional and derivational processing in older speakers. *Bilingualism: Language and Cognition*, 22(3), 425–440. https://doi.org/10.1017/S1366728918000615

Ricketts, J., Davies, R., Masterson, J., Stuart, M., & Duff, F. J. (2016). Evidence for semantic involvement in regular and exception word reading in emergent readers of English. *Journal of Experimental Child Psychology*, 150, 330–345. https://doi.org/10.1016/j.jecp.2016.05.013

Sabourin, L., & Haverkort, M. (2003). Neural substrates of representation and processing of a second language. In R. V. Hout, A. Hulk, F. Kuiken, & R. J. Towell (Eds.), *The Lexicon Syntax Interface in Second Language Acquisition* (pp. 175–195). Amsterdam: Benjamins.

Silva, R., & Clahsen, H. (2008). Morphologically complex words in L1 and L2 processing: Evidence from masked priming experiments in English. *Bilingualism: Language and Cognition*, 11(2), 245–260. https://doi.org/10.1017/S1366728908003404

Stuart, M., & Coltheart, M. (1988). Does reading develop in a sequence of stages? *Cognition*, 30(2), 139–181. https://doi.org/10.1016/0010-0277(88)90038-8

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. https://doi.org/10.1037/a0020126

Venezky, R. L. (1999). *The American way of spelling: The structure and origins of American English orthography*. NY: Guilford Press.

Weber, A., & Broersma, M. (2012). Spoken Word Recognition in Second Language Acquisition. In *The encyclopedia of applied linguistics* (Chapelle C ed., pp. 5368–5375). Wiley. https://doi.org/10.1002/9781405198431.wbeal1104

Wegener, S., Wang, H. C., de Lissa, P., Robidoux, S., Nation, K., & Castles, A. (2018). Children reading spoken words: interactions between vocabulary and orthographic expectancy. *Developmental Science*, 21(3), e12577. https://doi.org/10.1111/desc.12577

Wegener, S., Wang, H. C., Nation, K., & Castles, A. (2020). Tracking the evolution of orthographic expectancies over building visual experience. *Journal of Experimental Child Psychology*, 199, 104912. https://doi.org/10.1016/j.jecp.2020.104912

Wegener, S., Beyersmann, E., Wang, H.-C., & Castles, A. (2022a). Oral vocabulary knowledge and learning to read new words: A theoretical review. *Australian Journal of Learning Difficulties*, 27(2), 253–278. https://doi.org/10.1080/19404158.2022.2097717

Wegener, S., Wang, H.-C., Beyersmann, E., Nation, K., Colenbrander, D., & Castles, A. (2022b). Orthographic expectancies in the absence of contextual support. *Scientific Studies of Reading*, 1–11. https://doi.org/10.1080/10888438.2022.2127356

Appendix A List of Complex Novel Words

	Set 1		Set 2	
	complex words	stem morphemes	complex words	stem morphemes
Predictable	/dʒevɪŋ/ /dʒevd/ /dʒevz/	jev	/temɪŋ/ /temd/ /temz/	tem
	/jægɪŋ/ /jægd/ /jægz/	yag	/nɪdɪŋ/ /nɪdəd/ /nɪdz/	nid
	/vɪbɪŋ/ /vɪbd/ /vɪbz/	vib	/dʒɪtɪŋ/ /dʒɪtɪd/ /dʒɪts/	jit
	/tʌpɪŋ/ /tʌpt/ /tʌps/	tup	/jæbɪŋ/ /jæbd/ /jæbz/	yab
	/neʃɪŋ/ /neʃt/ /neʃɪz/	nesh	/vɪʃɪŋ/ /vɪʃt/ /vɪʃɪz/	vish
	/tʃɒbɪŋ/ /tʃɒbd/ /tʃɒbz/	chob	/ʃepɪŋ/ /ʃept/ /ʃeps/	shep
	/ʃʌgɪŋ/ /ʃʌgd/ /ʃʌgz/	shug	/θɒgɪŋ/ /θɒgd/ /θɒgz/	thog
	/θʌbɪŋ/ /θʌbd/ /θʌbz/	thub	/tʃɪgɪŋ/ /tʃɪgd/ /tʃɪgz/	chig
Unpredictable	/viːmɪŋ/ /vɪːmd/ /vɪːmz/	veme	/juːnɪŋ/ /juːnd/ /juːnz/	yune
	/baɪpɪŋ/ /baɪpt/ /baɪps/	bype	/kaɪvɪŋ/ /kaɪvd/ /kaɪvz/	kyve
	/jɜːpɪŋ/ /jɜːpt/ /jɜːps/	yirp	/bɜːvɪŋ/ /bɜːvd/ /bɜːvz/	birv
	/kɔɪbɪŋ/ /kɔɪbd/ /kɔɪbz/	koyb	/dʒaɪfɪŋ/ /dʒaɪft/ /dʒaɪfs/	jayf
	/dʒiːbɪŋ/ /dʒiːbd/ /dʒiːbz/	jeabb	/miːfɪŋ/ /miːft/ /miːfs/	meaph
	/fɜːfɪŋ/ /fɜːft/ /fɜːfs/	phirf	/gʌzlɪŋ/ /gʌzd/ /gʌzɪz/	ghuzz
	/gækɪŋ/ /gækt/ /gæks/	ghakk	/fegɪŋ/ /fegd/ /fegz/	phegg
	/mɜːbɪŋ/ /mɜːbd/ /mɜːbz/	mirbe	/veɪpɪŋ/ /veɪpt/ /veɪps/	vaype

Appendix B List of the Sentences Used in the Eye Tracking Experiment

	sentences
Set 1	Rick put his dirty socks into the machine to jev them clean.
	Diana put the best orange on the machine to veme the juice.
	Pam put the dirty flowers under the machine to yag them shiny.
	Max put his food into the machine to bype the green peas.
	Sara put her soaking wet hat on the machine to vib it dry.
	Lucy loaded all the rubbish into the machine to yirp it for recycling.
	Lucas put his sore tummy beside the machine to tup it better again.
	Jennifer put all her soggy chips under the machine to koyb them crispy.
	Nick put the playing cards into the machine to nesh before starting the game.
	Rex put the tennis balls into the machine to jeabb as he played fetch.
	James put the picture of the girl into the machine to chob her name.
	Jane put her cold and sore feet into the machine to phirf them warm.
	Matt put his feet into the machine so he could shug quickly up the wall.
	Sam saw a black bird and then made the machine ghakk to hear it sing.
	Ben put the machine into the fish tank to thub the dirty glass clean.
	Pip walked to the machine to mirbe his sandy body after sunbathing.
Set 2	Rick put his dirty socks into the machine to tem them clean.
	Diana put the best orange on the machine to yune the juice.
	Pam put the dirty flowers under the machine to nid them shiny.
	Max put his food into the machine to kyve the green peas.
	Sara put her soaking wet hat on the machine to jit it dry.
	Lucy loaded all the rubbish into the machine to birv it for recycling.
	Lucas put his sore tummy beside the machine to yab it better again.
	Jennifer put all her soggy chips under the machine to jayf them crispy.
	Nick put the playing cards into the machine to vish before starting the game.
	Rex put the tennis balls into the machine to meaph as he played fetch.
	James put the picture of the girl into the machine to shep her name.
	Jane put her cold and sore feet into the machine to ghuzz them warm.
	Matt put his feet into the machine so he could thog quickly up the wall.
	Sam saw a black bird and then made the machine phegg to hear it sing.
	Ben put the machine into the fish tank to chig the dirty glass clean.
	Pip walked to the machine to vaype his sandy body after sunbathing.

Note. From "Learning morphologically complex spoken words: Orthographic expectations of embedded stems are formed prior to print exposure," by Beyersmann et al., 2021, *Journal of Experimental Psychology: Learning, Memory, and Cognition, 47(1)*, pp. 97–98 (https://doi.org/10.1037/xlm0000808)