Northumbria Research Link

Citation: Bermúdez-Margaretto, Beatriz, Kopytin, Grigory, Myachykov, Andriy, Fu, Yang, Pokhoday, Mikhail and Shtyrov, Yury (2022) Biliteracy and acquisition of novel written words: the impact of phonological conflict between L1 and L2 scripts. Psychological Research, 86 (3). pp. 871-890. ISSN 0340-0727

Published by: Springer

URL: https://doi.org/10.1007/s00426-021-01529-y <https://doi.org/10.1007/s00426-021-01529-y>

This version was downloaded from Northumbria Research Link: http://nrl.northumbria.ac.uk/id/eprint/46325/

Northumbria University has developed Northumbria Research Link (NRL) to enable users to access the University's research output. Copyright © and moral rights for items on NRL are retained by the individual author(s) and/or other copyright owners. Single copies of full items can be reproduced, displayed or performed, and given to third parties in any format or medium for personal research or study, educational, or not-for-profit purposes without prior permission or charge, provided the authors, title and full bibliographic details are given, as well as a hyperlink and/or URL to the original metadata page. The content must not be changed in any way. Full items must not be sold commercially in any format or medium without formal permission of the copyright holder. The full policy is available online: http://nrl.northumbria.ac.uk/policies.html

This document may differ from the final, published version of the research and has been made available online in accordance with publisher policies. To read and/or cite from the published version of the research, please visit the publisher's website (a subscription may be required.)

Northumbria
University
NEWCASTLE

CUSTOMER SERVICE EXCELLENCE CSE

UniversityLibrary

Biliteracy and acquisition of novel written words: the impact of phonological conflict between L1 and L2 scripts

Beatriz Bermúdez-Margaretto[1*], Gregory Kopytin[1], Andriy Myachykov[1,2], Yang Fu[3], Mikhail Pokhoday[1] & Yury Shtyrov[1,4]

1 Centre for Cognition and Decision Making, Institute for Cognitive Neuroscience, National Research University Higher School of Economics, Russian Federation

2 Northumbria University, Newcastle, United Kingdom

3 University of La Laguna, Tenerife, Spain

4 Center of Functionally Integrative Neuroscience, Department of Clinical Medicine, Aarhus University, Denmark

***Corresponding author**: bermudezmargaretto@gmail.com

Abstract

The acquisition of new orthographic representations is a rapid and accurate process in proficient monolingual readers. The present study used biliterate and bialphabetic population to address the impact of phonological inconsistencies across the native (L1) and second (L2) alphabets. Naming latencies were collected from 50 Russian-English biliterates through a reading-aloud task with familiar and novel word forms repeated across 10 blocks. There were three Script conditions: (1) native Cyrillic, (2) non-native Roman, and (3) Ambiguous (with graphically identical, but phonologically inconsistent graphemes shared by both alphabets). Our analysis revealed the main effect of Script on both reading and orthographic learning: Naming latencies during training were longer for the ambiguous stimuli, particularly for the novel ones. Nonetheless, novel word forms in the ambiguous condition approached the latencies for the familiar words along the exposures, although this effect was faster in the phonologically consistent trials. Post-training tests revealed similarly successful performance patterns for previously familiar and newly trained forms, indicating successful rapid acquisition of the latter. Furthermore, we found the highest free recall rates for the ambiguous stimuli. Overall, our results indicate that phonological inconsistency initially interferes with the efficiency of novel word encoding. Nevertheless, it does not prevent efficient attribution of orthographic representations; instead, the knowledge of two distinct alphabets supports a more efficient learning and a better memory for ambiguous stimuli via enhancing their encoding and retrieval.

Keywords: reading, orthographic learning, biliteracy, Cyrillic alphabet, Russian language

Word count: 227

Orthographic learning refers to readers' ability to form novel representations of the word spellings in their mental lexicon (Share, 1995; 2008; Alvarez-Cañizo, Suarez-Coalla & Cuetos, 2019; Kwok & Ellis, 2015; Kwok, Cuetos, Avdyli & Ellis, 2017; Maloney et al., 2009). This ability develops when the reader is exposed to novel written word-forms, first learning how to read and later – during independent reading (Share, 1995; 2008). Thus, throughout repeated phonological decoding (i.e., print-to-sound) of novel words, new representations are formed for these stimuli enabling their direct visual recognition in subsequent encounters. This mechanism is crucial for the acquisition of novel vocabulary in the visual domain as well as for the development of efficient reading and communication skills.

Numerous studies have systematically investigated the near-immediate acquisition of novel written word-forms following very short training protocols of no more than ten exposures (Alvarez-Cañizo et al., 2019; Alvarez-Cañizo, Suárez-Coalla & Cuetos, 2018; Bowers, Davis, & Hanley, 2005; Clay, Bowers, Davis, & Hanley, 2007; Cunningham, Perry, Stanovich & Share, 2002; de Jong & Share, 2007; Kwok & Ellis, 2015; Kwok et al., 2017; Maloney et al., 2009; Martens & de Jong, 2008; Qiao, Forster, & Witzel, 2009; Qiao and Forster, 2013; Share, 1999; Shiffrin & Feustel, 1985; Suárez-Coalla, Álvarez-Cañizo, & Cuetos, 2016; Tamura, Castles & Nation, 2017; Wang, Castles & Nickels, 2012). Some of these studies have reported detectable learning effects already during training – for instance, in the reduction of the *lexicality effect*, i.e., the difference between naming latencies for novel and familiar words (e.g., Shiffrin & Feustel, 1985). Even more commonly, existing studies report the decrease of the length effect (i.e., the reduction of differences in naming latencies between short and long novel words) with learning (e.g., Alvarez-Cañizo et al., 2019; Kwok & Ellis, 2015; Kwok et al., 2017; Maloney et al., 2009). These results indicate a change in the reading strategy for the trained words as a consequence of the formation of orthographic representations evolving from sequential letter-by-letter decoding to a whole-word visual recognition strategy.

Other studies provide evidence regarding post-training access to newly-acquired orthographic representations, manifested as better recall (through typing) or better recognition (among novel non-trained foils) of the newly trained words (Cunningham et al., 2002; Share, 1999). Similarly, some studies register interference effects caused by the learned stimuli during processing of orthographically related familiar words, which is taken as a sign of lexical competition following the integration of novel items into the mental lexicon (Bowers et al., 2005; Clay et al., 2007; Qiao et al., 2009; Qiao & Forster, 2013). In general, these and similar findings indicate the emergence of very rapid and robust effects of orthographic learning in different languages such as Spanish (Alvarez-Cañizo et al., 2018; Suárez-Coalla et al., 2016), Italian (Paulesu et al., 2000), English (Tamura et al., 2017; Wang et al., 2012), and Dutch (Martens & de Jong, 2008; de Jong & Share, 2007).

Importantly, the majority of studies reporting rapid orthographic learning use monolingual populations, thus exploring the acquisition of novel words in L1 reading. However, due to the global growth of bilingual population, the number of proficient second language (L2) *readers* (so called *biliterates*) who incorporate new vocabulary through L2 reading is constantly growing. Moreover, biliteracy often implies managing a typologically different alphabet or *script* (in which the same unit of the spoken language, such as the phoneme, is represented by different written characters or graphic signs, as, e.g., in Latin, Greek or Cyrillic scripts) or even a different writing system (in which not only the scripts are different, but also the units of the spoken language that are represented, as in English or Japanese Kana, where written characters represent phonemes or syllables, respectively).

These cases are quite common in the highly multilingual world, and they motivate several specific research questions. First, does orthographic learning of L1 and L2 words differ, especially in situations when the two writing systems and/or scripts are substantially different? Second, what are the factors that influence orthographic learning in such cases? The past few years have seen numerous reports documenting systematic cross-lingual transfer effects between L1 and L2 (see Lallier & Carreiras, 2018, and Chung, Chen, & Geva, 2019 for recent reviews). For instance, several studies have addressed the impact of mapping the same letters onto different sounds across languages, by studying bilinguals whose two languages share the same alphabet (e.g., Brysbaert et al., 1999; Jared & Kroll, 2001; Jared & Szucs, 2002; von Studnitz & Green, 2002). Importantly, these studies show that bilinguals activate phonological representations from both of their languages simultaneously. Other reports examining biliterates across different writing systems or scripts, such as Japanese-English (Ando, Jared, Nakayama & Hino, 2014; Nakayama, Sears, Hino & Lupker, 2012), Korean-English (Kim & Davis, 2003), Chinese-English (Zhou, Chen, Yang, & Dunlap, 2010) or Russian-English (Jouralev, Lupker & Jared, 2014) report cross-script phonological priming effects. In particular, phonological representations generated from primes in one script facilitate the recognition of targets in another script, thus suggesting integration of phonological representations across languages. However, although extensive research has studied reading across different languages and scripts, fewer studies have addressed the interplay between two writing systems or scripts during orthographic *learning*. The present study presents an attempt to fill this gap by studying orthographic learning in a biliterate population fluent in two different languages which employ two different scripts.

Reading is often conceptualized as extracting or decoding phonological information from visual input by translating written messages into spoken ones (Perfetti, 2003; Share & Stanovich, 1997). This is particularly true for unknown or low-frequency words whereby a sublexical translation of print-to-sound must be carried out in the absence of any pre-existing phonological or semantic representations. For high-frequency words, however, the access to whole-word orthographic representations also enables reading via a direct access to the phonological or the semantic levels of word-specific information, thus effectively allowing a double-route access (Grainger & Ziegler, 2011). In general, reading is likely to be rather similar across L1 and L2 orthographies, since the core process is universal across writing systems. However, the existence of a variety of scripts (within and across different writing systems) as well as their variations lead to incongruencies between the native and non-native scripts. This heterogeneity affects both decoding and subsequent word learning in the L2 alphabet. In particular, writing systems and scripts may differ in terms of the *visual complexity* of the graphemic representations: Some contain a small number of strokes and a relatively small set of written characters (e.g., Spanish or Hebrew) while others include a much wider variety of strokes and a larger set of written characters (e.g., Arabic, Indian Kannada, or Chinese) (Nag, 2007; McBridge-Chang, Zhou, Cho, Aram, Levin & Tolchinsky, 2011; Abdelhadi, Ibrahim & Eviatar, 2011). Alphabets may also differ in their *orthographic representation:* the grain size, or unit of spoken language that is represented by visual characters, from single phonemes, like in French or Russian, to syllables or morphemes, like in Japanese Kana or Chinese, respectively (Perfetti, Liu, & Tan, 2002; Perfetti & Dunlap, 2008). Finally, it is customary to refer to the *orthographic depth* or the transparency/regularity in the print-to-sound mappings. Orthographic depth ranges from consistent one-to-one mapping of graphic signs onto sounds in transparent orthographies like Finnish, Greek or Indian Devanagari to inconsistent one-to-many and many-to-one mappings in opaque orthographic systems like in English or French (Seymour, Aro & Squirne, 2009; Ziegler & Goswami, 2005).

Consequently, efficient decoding and orthographic learning in an L2 script depend on several challenging tasks. First, biliterates must learn the orthographic units specific to L2 script as well as their phonological correspondences. Second, they need to manage the eventual inconsistencies between L1 and L2 scripts, meaning that a lower level of incongruency allows for better decoding and orthographic learning outcomes. Thus, orthographic inconsistency between L1 and L2 scripts, in terms of the level of orthographic representation or orthographic regularity, leads to different reading strategies affecting word learning in the non-native script (Hamada & Koda, 2008; Schwartz et al., 2014). For instance, a study investigating novel word learning in Korean-English and Chinese-English biliterates showed that the decoding and recall of novel words in L2 English was better among Korean learners of English. This effect is explained by a higher consistency between Korean and English reading, since both follow phonological decoding of visual patterns (letters) into sounds, whereas in Chinese, graphic signs (logographs) are holistically transformed into a whole morpheme (Hamada & Koda, 2008).

Furthermore, the level of phonological inconsistency across alphabets, in terms of the symbol-to-sound correspondences carried out in L1 and L2 reading, could also play an important role both in the acquisition of decoding and orthographic learning skills in the L2 alphabet or script. However, the effects of phonological inconsistencies across L1 and L2 scripts, led by decoding of the same visual characters into different sounds across scripts, have been poorly understood so far, likely due to the lack of research into cross-linguistic grapheme overlap. Nonetheless, several studies have addressed the effect of orthographic overlap across languages with the same alphabet by means of cognates –words that are translation equivalents and have total or partial orthographic overlap across L1 and L2 (such as ‘’piano’’ or ‘’tomaat’’, examples of identical and non-identical English-Dutch cognates; Bultena, Dijkstra & van Hell, 2013; Cop, Dirix, Van Assche, Drieghe, & Duyck 2017; Peeters, Dijkstra & Grainger, 2013). In these studies, the grapheme overlap typically leads to a faster processing of cognates than non-cognate control words, reflecting a cross-lingual facilitatory effect resulting from the coactivation of both words within an integrated lexicon. However, as orthographic cognates usually share aspects of phonological decoding (i.e., the cognate word *piano* is pronounced similarly in English and Dutch), such effects may also be explained by phonological rather than purely orthographic factors, and they can be rather different when the overlap involves phonological inconsistencies across languages and scripts. This view is supported by the studies exploring the impact of the phonological incongruency across L1 and L2 scripts using visual word recognition tasks (Havelka & Rastle, 2005; Lukatela, Lukatela, Carello, & Turvey, 1999; Lukatela & Turvey, 1990; Rastle, Havelka, Wydell, Coltheart & Besner, 2009). In these studies, mainly conducted in Serbo-Croatian - English biliterates, L2 English words are systematically named *slower* when they contain ambiguous or inconsistent graphemes that sound differently in L1, than those containing non-ambiguous, L2-specific graphemes. This phonological ambiguity effect is generally attributed to the application of two competing decoding rules to the graphemes with the same visual but different phonological representations in the two languages.

Importantly, Cyrillic and Latin (Roman) alphabets are direct descendants of the ancient Greek alphabet and are characterized by having a considerable degree of graphemic overlap. Cyrillic script is used by a relatively large number of readers in the world across many Slavic (such as Russian, Belarusian, Ukrainian, Serbo-Croatian, Bulgarian, or Macedonian) and non-Slavic languages (such as Tatar, Mongolian, and Ossetic). At the same time, users of Cyrillic are usually extensively exposed to the Roman alphabet through education, literature, TV, cinema, advertising etc. Moreover, this population is also often relatively proficient in English and/or other Western European languages, and thus fluent in reading Roman script. In addition, some cultures, such as Serbian, use Roman and Cyrillic scripts interchangeably in their native language. Although both Roman and Cyrillic alphabets have their script-specific or unique graphemes (i.e., ш, ж, ф, ч only present in Cyrillic; v, q, z, f only present in Roman), several graphemes are used in both scripts. Some of these shared graphemes are mapped onto a comparable phonological representation (i.e., k, t, o, a), whereas others have a different mapping across scripts (i.e., “p”, decoded as /p/ in Roman but as /r/ in Cyrillic, or “н”, decoded as /h/ in Roman but as /n/ in Cyrillic). As a result, words sharing graphemes in both scripts have different phonological mapping and meaning across both languages (i.e., “*cop”*is read as /cop/ in Roman but as /sor/ in Cyrillic, which means “litter” in Russian). Although such phonological ambiguity effects have been showed to affect the reading latencies in Cyrillic-Roman biliterates, it is unclear how the corresponding processes unfold during novel word learning. Taking into account the key role of phonological assembly in reading, especially in alphabetic languages such as English (Ehri, 1992; Goswami & Bryant, 1990; Share, 2008; Snowling & Göbel, 2011), it is reasonable to expect that inconsistent phonology across alphabets would impact the formation of new L2 orthographic representations through phonological decoding processes, and hence the acquisition of efficient L2 reading skills.

The present study explores the orthographic learning processes in a group of Russian-English biliterates. Here, we disentangled the effect of phonological inconsistency from the effect of novel word learning *per se* by manipulating consistently the L2 alphabet stimuli. Namely, we used the L2 stimuli consistent or inconsistent with the L1 decoding rules. The novel word forms were repeatedly presented in different script conditions (Cyrillic, Roman, ambiguous) in a reading-aloud task together with familiar words enabling us to test the impact of the training on lexical differences. We formulated two specific research questions. First, we aimed to determine whether L1-L2 phonological inconsistencies influence the formation of new orthographic representations. Second, we investigated whether such inconsistencies are also reflected in the subsequent access to these representations (through active and controlled retrieval and through automatic recognition). To these ends, we measured orthographic learning process online during the training phase (through a reading-aloud task) and offline –

at the post-training session – via recall, recognition, and lexical decision tasks. Recall and recognition tasks were used to directly evaluate the access to newly-represented word forms – either via active and controlled retrieval (recall task) or via an automatic and superficial familiarity-based process (recognition task). Lexical decision task was used to evaluate the orthographic representations of novel word forms indirectly – by examining the corresponding interference during a forced stimuli categorization. Following other word learning studies using similar approaches (Merkx, Rastle, & Davis, 2011; Leminen, Kimppa, Leminen, Lehtonen, Mäkelä & Shtyrov, 2016), we hypothesized that the newly trained novel word forms would be more difficult to reject as non-words than completely unfamiliar pseudowords, which would be reflected in differential response latencies and/or accuracy. Regarding our first research question, we hypothesized that phonological inconsistencies related to the L1-L2 overlap would affect the formation of new orthographic representations since such inconsistencies should lead to an interference in the phonological decoding during orthographic learning. Therefore, we expected that access to the novel L2 words with inconsistent graphemes would be associated with longer naming latencies as well as with a smaller reduction of lexical differences with familiar words, in comparison with novel L2 words with graphemes consistently decoded across L1 and L2 alphabets. As for our second research question, we expected that new orthographic representations acquired under phonological ambiguity would be poorly accessed both in recall and recognition tasks during the post-training assessment. Moreover, we expected a poor access to the novel L2 words with inconsistent graphemes reflected in the lexical decision task performance. Particularly, we expected a more frequent categorization of these stimuli as non-words to cause a weaker interference than the categorization of the words well represented in reader's lexicon showing longer latencies and higher number of errors. Finally, a general effect of native alphabet was expected, with better orthographic representation and access for novel words presented in L1 Cyrillic script.

Materials and Methods

Participants

Fifty students recruited from the National Research University - Higher School of Economics and Moscow State University of Psychology & Education (23 females, aged between 18 and 30 years old, M_{age} = 20.8, *SD* = 2.78) took part in the experiment. All participants were right-handed native Russian (L1) speakers with normal or corrected-to-normal vision and no history of cognitive, neurological, or psychiatric disorders. All of them had English as their second language (L2), with different speaking and reading proficiency levels and relatively late learning onset (see Table 1 for details). In addition, 34 of them were also speakers of other languages (L3)[1]. The study was approved by the Ethics Committee of the Department of Psychology, National Research University Higher School of Economics.

Table 1 here

Stimuli

Experimental stimuli consisted of 12 familiar words and 12 pseudowords (namely, orthographically legal but meaningless stimuli, thus acting as novel word forms to be learned). These stimuli were equally divided into unambiguous L1 Cyrillic, unambiguous L2 Roman, and ambiguous script conditions. Therefore, 12 familiar words were presented in (4) Cyrillic (e.g., "*шаг"*), in (4) Roman, (e.g., "*vet",*) or in (4) ambiguous script (e.g., "*cop")* and the same was done for the 12 novel word forms (4 in Cyrillic, e.g., *"шаз";* 4 in Roman, e.g., *"vaz";* and 4 in ambiguous script, e.g., *"pex"*). The full list of experimental stimuli can be found in Appendix 1. All stimuli were 3 letters in length with a Consonant-Vowel-Consonant (CVC)

---

[1] German: n=19; French: n=11; Spanish: n=6; Latin n=4; Italian: n=3; Ukrainian: n=3; Arab: n=2; Armenian: n=1; Chinese: n=1; Swedish: n=1; Indonesian: n=1; Czech: n=1; Belorussian: n=1; Danish: n=1

structure. Novel words were designed maintaining the first letter of a familiar word in the corresponding script condition to ensure maximal similarity. In addition, stimuli presented in L1, L2 and ambiguous conditions were matched across each group of familiar and novel words in log trigram frequency (paired tests carried out using nonparametric Mann-Whitney-Wilcoxon confirmed no differences across conditions, all contrasts *p*>.1). Trigram frequency values for L1 and L2 stimuli were taken from Russian National corpus (http://www.ruscorpora.ru/new/search-main.html) and British National Corpus (https://www.english-corpora.org/bnc/) online databases, respectively, and log transformation was applied in order to normalize both datasets.

Importantly, all unambiguous stimuli in Cyrillic and Roman scripts were designed by using graphemes specific to each alphabet (e.g., *j, ш*) as well as those common in both languages and mapped onto the same phonemes (*a, м)*, i.e., phonologically consistent across scripts. However, in the ambiguous condition, stimuli were created by combining common and consistent graphemes with common but inconsistent graphemes, namely those used in both Cyrillic and Roman alphabets but decoded into a different sound depending on the script (i.e., the grapheme “p” is decoded as /p/ in Roman but as /r/ in Cyrillic, and the grapheme “x” is decoded as /ks/ in Roman but as /h/ in Cyrillic). In order to ensure the stimuli ambiguity in the ambiguous condition, handwriting fonts were used in the study since they provide a larger choice of overlapping graphemes (see Appendix 1). For instance, the English word “*cop*” written this way reads as /sor/ in Russian (meaning “litter”), thus resulting in the decoding ambiguity. The combination of italic “Notperfect regular” and “Swanky and Moo Moo Cyrillic” fonts with small manual edits was used to optimize the letters for these purposes. The same handwriting style font was used for both the training and post-training tasks, including the corresponding experimental instructions. An additional stimuli set was used as foils for the recognition and lexical decision tasks during the post-training phase. Thus, for each task, 48 untrained stimuli were constructed (2 foils per each previously trained stimulus), maintaining the first two letters of the corresponding stimulus in the training task and replacing the third one to make a novel foil item, ensuring sufficient difficulty of the post-training assessment tasks.

Procedure

The duration of the entire experiment was approximately one hour. Participants underwent (1) a training phase consisting of a reading-aloud task and (2) a post-training test phase comprising recall, recognition, and lexical decision tasks to assess the outcomes of learning. Before the training phase, participants also completed the full version of the Leap Questionnaire (Marian, Blumenfeld & Kaushanskaya, 2007), a tool that collects self-reported, subjective judgements of L2 proficiency and L2 exposure, coded in 0-100 and 0-10 scales, respectively. A minimum level of L2 proficiency (at least 15 points on the general scale for L2 proficiency) was required for participation in the study, ensuring low-to-medium L2 proficiency and capability for L2 reading.

During the training phase, participants were presented with the set of 24 familiar and novel word forms repeatedly across 10 different blocks (see Figure 1 for experimental sequence). The participants were told they would see a series of both familiar and novel words, presented either in L1 (Russian) or in L2 (English) and they were asked to read them aloud as quickly and accurately as possible. The stimuli were presented in black font against a grey background at the centre of a computer screen by means of E-Prime software (Psychology Software Tools, Inc., Schneider, Eschman & Zuccolotto, 2002). A Microsoft LifeChat LX-3000 headset (with a noise-cancelling microphone) was used to collect participant's vocalisations for each stimulus. Stimulus presentation was pseudorandomized within each block and participant in order to prevent the presentation of two consecutive stimuli from the same condition. Moreover, given that the previous presentation of an L1 or L2 stimuli could bias the pronunciation of an ambiguous stimulus, each trial included a distractor target stimulus (a white or a black diamond) presented between target stimuli (see Figure 1). Participants had to indicate the color of the diamond by pressing the corresponding keyboard key with their right (L) or their left finger (D). Similar to previous studies addressing cross-language naming (e.g., Reverberi et al., 2018), such an inter-trial non-linguistic distractor task (the categorization of the target color) was introduced to prompt participants to disengage from the reading processes thus preventing or minimizing the influence of the preceding stimulus script on the language chosen to read the ambiguous stimuli. The keys were labeled with corresponding color stickers. The color of distractor stimuli was randomized across trials and responses were counterbalanced across participants (namely, half of them responded to white color with their right index finger and black with their left index finger whereas the other half did the opposite). Before starting the training task, participants were presented with 12 practice trials (2 trials per condition) using stimuli which were similar, but not identical to the main task. During the training, participants took two breaks (after 4th and 7th blocks) in order to avoid fatigue.

Figure 1 here

Immediately after completing the reading-aloud task, participants underwent the post-training phase starting with a recall task, in which they were asked to write down all stimuli they could remember from the previous training phase. Answers were collected on a paper sheet with 30 spaces to fill, with no time restriction. Immediately after that, participants carried out the recognition and the lexical decision tasks sequentially with the same procedure and stimuli, but with different instructions. In both, the stimuli previously presented during the training phase were presented together with foils. Stimuli were presented in randomized order at the centre of a computer screen by means of E-prime software. Participants were asked to press a button on a keyboard (D or L, labeled with a white or a black sticker) to decide whether the stimulus had been previously presented in the training phase or not (in the recognition task) and to categorize the stimulus as a real word or non-word (in the lexical decision task). For half of the participants, D button was labeled with a white sticker and L button with a black sticker, with the opposite for the other half; moreover, half of the participants pressed white for those stimuli previously trained and black for the stimuli not previously trained, whereas the other half did the opposite. Color coding rather than characters (such as Y/N) were used to avoid any idiosyncratic influence of reading key labels on the main task. Response time was not limited; both latency and accuracy were collected in both tasks.

Data Analysis

Training phase

Reading latencies obtained at the reading aloud task were extracted manually for each trial and participant using Praat software (Boersma, 2011). Utterances containing errors (incorrect pronunciations, mix of alphabets in ambiguous stimuli, no response or hesitation sounds, in which voice sounds, such as “err...’, ‘uhm...’ etc. are produced but the stimulus is not named) were excluded from the analysis (representing 2.33% of all trials). Therefore, responses to ambiguous stimuli were considered equally correct if read in Russian or English, but not if the utterance included a mixture of both alphabets. In addition, responses whose latencies were 2 standard deviations above or below the mean were also rejected (4.16% of data). The ambiguous nonword stimulus “*сик*” was also excluded from the analyses because this stimulus was pronounced in Russian in 99.6% of cases, thus being an outlier in the ambiguous subset.

In order to determine the effect of the presentation script on the reading latencies of both novel and familiar words across the training blocks, an inferential analysis using mixed-effects modelling was carried out. This method allows to simultaneously enter random participant or item effects in addition to the experimental variables thus effectively separating the fixed effects of predictor variables (Barr, Levy, Scheepers, & Tily, 2013; Baayen, Davidson & Bates, 2008). The analysis was conducted in R software (Team, 2013) using the lmer package (Baayen et al., 2008). Block (from 1 to 10), Lexicality (familiar and novel words), and Script (Cyrillic, Roman and Ambiguous) were entered as predictor variables (fixed effects), participants and items were treated as random effects, and the RTs – as the dependent variable. The final model included fixed effects for lexicality, script and block, random intercepts for participant and item and by-participant random slopes for lexicality and script. More complex model with all the within-subject and within-item predictors as random slopes failed to reliably converge (Barr et al., 2013). Variance inflation factor (ranged from 1.01 to 1.32) reported no collinearity issue in the model. R-default treatment contrasts were altered to sum-to-zero contrasts before running the model, so that all fixed effect were contrast-coded (Schad, Vasishth, Hohenstein, & Kliegl, 2020). We used lmerTest packages (Kuznetsova, Brockhoff & Christensen, 2017) to calculate *p*-values and Type III *F*-statistics for main effects and interactions using Satterthwaite approximations to determine degrees of freedom. Post-hoc analysis was achieved using the framework provided by the emmeans package (Lenth, 2018). We first determined the estimated marginal means (EMMs) and their standard errors, and then pair-wise comparisons. In addition, we used asymptotic *dfs* (i.e., *z* values and tests) to prevent emmeans from calculating the *df* for the EMMs.

We further conducted a Bayes Factor analysis in order to adjust P values (BF01 < .0001), to quantify the statistical evidence supporting the interaction between the three factors. The Bayes factor analysis was calculated using the Bayesian Information Criterion (BIC) approximation of the Bayes Factor (Wagenmakers, 2007). The Bayes Factor BF01 was calculated using the BIC values for the model without the interaction (the null hypothesis H0) and for the model with the interaction (the alternative hypothesis H1), using the formula BF01 = exp((BIC(H1) – BIC(H0))/2) (Wagenmakers, 2007, p. 796). A BF01 less than 1 would suggest evidence in support of H1 (i.e., the alternative hypothesis), whereas BF01 greater than 1 would suggest evidence in support of H0 (i.e., the null hypothesis) and BF01 = 1 would suggest equivalent evidence for the two hypotheses.

Complementary analyses were carried out in order to determine the language in which stimuli presented in the ambiguous script were read across the training blocks (see supplementary material section A). In addition, linear regression analyses were also implemented to study the effects of (1) L2 proficiency level, (2) L2 age of acquisition, and (3) exposure to L2 on the training effectiveness reflected in the naming latencies obtained at the end of the exposures (see supplementary material section B).

Post-training phase.

For the recall task, the number of correct recalled stimuli across each condition and participant was calculated and converted into to a percentage scale. A two-way rmANOVA (using R function aov) was performed taking the percentage of correct responses as a dependent variable, as well as lexicality and script as independent variables. All *p* values were adjusted using the FDR correction method for multiple comparisons. For recognition and lexical decision tasks, only correct responses were entered in the analysis (discarding 8.36% of trials in recognition and 13.07% in the lexical decision task). All responses below 500 ms and above 2500 ms were excluded from the analysis (excluding 5.64% and 15.93% of responses in recognition and in lexical decision task, respectively). Then, the remaining RTs below or above 2 standard deviations were excluded from data analysis. Accuracy data and RTs of correct responses were analyzed by using generalized and linear mixed-effect models, respectively. All models were conducted with familiarity (trained, non-trained foil), script (Russian, English, Ambiguous) and lexicality (familiar, novel) as fixed effects and item and participant as random effects. The final models of both recognition and lexical decision task included by-participant random slopes for lexicality and by-item random slopes for lexicality. More complex model with all relevant random structures failed to reliably converge. Car package with the Type III Wald $\chi 2$-statistics (function Anova) (Fox & Weisberg, 2016) was used to test for significance for accuracy data and to calculate *p* values.

Results

Training Phase

Table 2 here

Table 2 shows the mean latencies obtained in the reading aloud task, separated by block (first vs. last), lexicality (familiar vs. novel) and script type (L1, L2, ambiguous). The final model revealed statistically reliable main effects of *block*, as naming latencies significantly decrease across the training (Block 1: *M*=832; Block 10: *M*=634), *lexicality*, given novel words showed longer naming latencies (*M*=738) than familiar words (*M*=663) and *script*, with naming latencies differing depending on the alphabet of presentation (L1: *M*=606; L2: *M*=711; Ambiguous: *M*=784). We also found reliable interactions between *block x script*, *block x lexicality* and marginally *lexicality x script*; importantly, the three-way interaction *block x script x lexicality* was also found reliable. See Table 3 for detailed statistical results. A Bayes Factor analysis was conducted to evaluate the statistical evidence for the three-way interaction. The Bayes Factor BF01 was calculated using the BIC values for the model with no interaction (the null hypothesis H0) and for the model with an interaction of block by lexicality by script (the alternative hypothesis H1). The Bayes Factor, BF01 <0.001, indicates "strong" evidence for the alternative hypothesis (i.e., an interaction between three predictors) according to Jeffreys's (1961) classification scheme.

Moreover, a power analysis was conducted using *simulate ()* function in the lmer4 package (Bates et al., 2015) within the R statistical computing environment. This study includes a relatively small number of items per condition, which was necessary in order to fulfil the strict requirements controlling various orthographic and psycholinguistic variables across conditions. To ensure no lack of power in our results, and based on the mixed-effect model that included the same effect structure (fixed effects for script, block and lexicality, random intercepts for participant and item and by-participant random slopes for script and lexicality), 1000 simulations of the model were conducted separately for each condition as well as two-way and three-way interactions. It produced an estimate of statistical power of .965 with our sample of 50 participants (i.e., in 965 out of 1000 simulation runs, the model detected a significant three-way interaction of block by script by lexicality), thus largely ruling out the possibility that the present results may lack statistical power.

Table 3 here

The three-way *block x lexicality x script* interaction confirmed that repeated exposures across the ten training blocks led to a decrease in the differences between familiar and novel words, although differently depending on the script of presentation. In order to disentangle the three-way interaction, the interaction between *lexicality x script* across the training blocks was explored (see Table 3). Post-hoc analyses (pairwise comparisons) revealed that in the first training block, familiar and novel words were differently influenced by the script effect, with novel words showing longer naming latencies when read in L2 script, and, particularly, in the ambiguous script, in comparison to those read in L1. However, familiar words showed similar effect of non-native alphabet when presented in L2 or in ambiguous script in comparison to L1. Nonetheless, in the last training block, novel and familiar stimuli did not any more differ in their naming latencies depending on the script of presentation. Hence this pattern of results indicates a different impact of the script effect for novel and familiar words which changed across the training. Figure 2 shows the reading latencies pattern across the training blocks for all familiar and novel words.

Figure 2

Second, the *lexicality x block* interaction was explored across the levels of script factor. These results confirmed that differences between novel and familiar words were differently reduced along the exposure as a function of the script, due to different effects of training on each of the scripts (see Table 3). In particular, pos-hoc comparisons for the effect of block revealed that, for familiar words, the highest latency decrease across the ten exposures was found in L2 followed by ambiguous and by L1 scripts. However, for novel word forms, the highest reduction in naming latencies was registered for the stimuli presented in ambiguous and L2 scripts followed by those presented in L1 script. Such different pattern in the naming latency drop for novel and familiar words led to a different attenuation of the lexicality effect across scripts (see Table 4 for detailed statistical results across all blocks). Although, in general, initial differences for familiar and novel stimuli in the beginning of the training were found reduced at the end of the task, this reduction was faster in non-ambiguous L1 and L2 conditions than in the ambiguous script condition. Thus, the initially reliable differences between familiar and novel words presented in both L1 and L2 scripts were eliminated already by the third and second presentation, respectively. However, reliable differences between familiar and novel words presented in ambiguous script disappeared only at the last presentation.

Table 4

Post-training Phase

Recall task

The rmANOVA revealed a statistically reliable effect of *script* (*F*(2,98)=22.464, *p*<.001). Pair-wise comparisons, using paired *t*-test, revealed differences across the three alphabets, with better recall rates for stimuli presented in the ambiguous script (59.4%) than those presented in L2 (47.7%, *p*<.0001) or L1 alphabets (36.5%, *p*<.0001); moreover, recall scores also differed between L1 and L2 scripts, with the lowest percentage of correctly recalled stimuli found in the native alphabet (*p*<.0001).

Figure 3

Recognition task

The final model revealed reliable effects of *familiarity*, showing lower recognition latencies (i.e., faster recognition) for trained (*M*=923 ms) than for untrained stimuli (*M*=1008 ms) and *lexicality*, with lower recognition latencies for familiar (*M*=919 ms) than for novel words (*M*=1012 ms), regardless of the script of presentation (see Table 5). The interaction *familiarity x lexicality* was marginally significant, suggesting that while familiar words showed similar recognition latencies independently of previous training, novel trained word forms were recognized significantly faster than untrained ones. Thus, novel words exhibited similar recognition times to familiar words when both had been trained, whereas untrained stimuli showed reliable lexical differences. Moreover, the interaction s*cript x familiarity* was also found statistically significant, indicating that stimuli presented in L1 or in ambiguous script showed similar recognition latencies independently of a previous training, whereas those presented in L2 exhibited significantly faster recognition latencies when trained than when presented for the first time in the recognition task. Previously trained stimuli showed similar recognition times regardless of the script of presentation whereas for untrained stimuli, those presented in L1 were discarded slightly faster than L2, although no differences were observed between L1 and ambiguous or ambiguous and L2 foils. See Figure 3B. As recognition accuracy was close to 100% across all stimuli, no further analyses were carried out on this data.

Table 5

Lexical decision task

The analysis carried out on response latencies revealed main effects of *script*, indicating longer reaction times for stimuli presented in ambiguous (*M*=1114 ms) and L2 (*M*=1138 ms) than in L1 alphabet (*M*=963 ms) and *lexicality*, revealing generally longer response latencies for novel (*M*=1227 ms) than for familiar words (*M*=929 ms). We also found reliable interactions of *script* × *lexicality* and *familiarity* × *lexicality*. No other reliable effects or interactions were registered. Post-hoc analyses were carried out in order to explore these interactions in detail (see Table 6). Regarding *script x lexicality* interaction, we found that novel words showed longer response times when presented in L2 and, particularly, when presented in ambiguous script, in comparison to those presented in L1 alphabet; in contrast, the categorization of familiar words was particularly delayed when presented in L2 than in ambiguous or in L1 script. The highest RT differences for the lexical categorization of novel and familiar words were found in the ambiguous condition, although also significant in L1 and L2. Regarding the interaction *familiarity x lexicality*, it revealed that familiar words showed faster reactions for previously trained than for untrained foils, whereas novel words exhibited the contrary pattern, with significantly longer RTs for trained than for foils, thus reflecting an interference effect when categorizing these stimuli as non-lexical items. As a result, differences between novel and familiar words were larger when these stimuli were previously trained than when stimuli received no previous training (See Figure 3C).

Table 6

The relatively high error rate in this task (13.07% of the total amount of responses), also allowed for LME analyses to be conducted on accuracy data. The final model revealed a main effect of *script*, with higher percentage of correct categorization of stimuli in L1 (*M*=97.25 %) than in L2 (*M*=82.25 %) or in ambiguous script (*M*=86.8 %). In addition, reliable interactions were found between *familiarity x lexicality* and *script x lexicality* (see Table 7 for detailed statistical results).

Table 7

Post-hoc analyses were carried out to further investigate the interactions. Regarding *lexicality x script* interaction, these revealed significantly better categorization of familiar than novel words when presented in ambiguous script, whereas no differences were observed between these stimuli when presented in L1 or L2 (see Figure 3D). Thus, familiar words in ambiguous and L1 scripts showed more correct responses than those presented in L2, whereas no differences were observed between ambiguous and L1 script. However, novel word forms showed a different pattern in accuracy, as the greatest percentage of correct responses (i.e., rejections) was observed for those presented in L1 in comparison to those in L2 or in ambiguous script. No accuracy differences were observed between the categorization of novel words in L2 and ambiguous script. Regarding the analysis of the *lexicality x familiarity* interaction, it was found that familiar words used in the training block exhibited significantly better categorization than those untrained; the opposite pattern was obtained for novel word forms, with worse categorization performance observed for the trained items than for foils, reflecting an interference during the categorization of trained novel words as non-lexical stimuli. Therefore, novel words exhibited significantly lower percentage of correct categorization responses than familiar words when these stimuli were previously trained but better categorization than familiar words when presented as foils (see Figure 3D).

Discussion

The present study investigated orthographic learning of novel words during the use of native and non-native alphabets. In particular, we aimed to determine the impact of phonological inconsistencies established as graphemic overlap between L1 and L2 scripts. To this end, we administered training and post-training tasks in order to evaluate the processes of orthographic representation building and subsequent access to novel written word-forms in a group of Russian-English biliterates, who were exposed to familiar and novel words in a short training session containing visually presented words in L1 Cyrillic, L2 Roman or ambiguous scripts. Overall, our results demonstrate that phonological inconsistency interfered with the reading automatization of novel word-forms. This interference slowed down but did not prevent the formation of orthographic representations since similar levels of post-training recall and recognition were observed for the trained and for the familiar words even in the conditions of phonological inconsistency. We discuss these results in more detail below.

The short ten-exposure training protocol used in this study led to a rapid automatization in the decoding of novel word-forms presented in either L1 or L2 alphabets. This result corroborates previous findings using similar training protocols and online measures (Alvarez-Cañizo, Suarez-Coalla & Cuetos, 2019; Kwok & Ellis, 2015; Kwok, Cuetos, Avdyli & Ellis, 2017; Maloney et al., 2009; Shiffrin & Feustel, 1985). Specifically, naming latencies progressively decreased to the extent that they matched those exhibited by familiar words at the end of the training. Nonetheless, the improvement in the naming latencies and acquisition of orthographic representation depended on the phonological ambiguity of the stimuli's grapheme-to-phoneme correspondence. Thus, the reading of novel words with ambiguous graphemes was slower than reading of the words containing unambiguous graphemes. However, this ambiguity effect was mainly present at the beginning of the training when these stimuli were still unfamiliar, but it was significantly reduced at the end of the training. This pattern suggests that the orthographic familiarity established through repeated exposure mitigated the impact of phonological inconsistency. Similarly, the impact of the phonological ambiguity was stronger for novel words (cf. Havelka et al., 2005; Lukatela et al., 1999) but, importantly, this effect was registered at the beginning of the training only, indicating that the decoding of novel stimuli at the early training stages is carried out in a serial manner via grapheme-to-phoneme correspondence, and is therefore particularly affected by the ambiguity. However, as new stimuli gained orthographic familiarity across the training and led to the incorporation of new representations in the reader´s lexicon, the impact of phonological ambiguity decreased. Indeed, the reading of both novel and familiar ambiguous words was similarly affected by phonological inconsistencies at the end of the training, indicating that by that point the novel words followed a reading strategy less dependent on phonological decoding and more on the direct access to their newly-acquired memory traces.

Importantly, acquisition of orthographic representations for novel words with ambiguous graphemics was registered both in online and offline measures. First, it was observed as the elimination of the lexicality effect in the ambiguous condition at the end of the training, an effect that typically reflects differences between familiar lexicalized words and new unfamiliar entries (e.g., Forster & Chambers, 1973). This finding likely indicates the incorporation of the newly-trained ambiguous L2 word forms into the orthographic lexicon regardless of the inconsistencies in their decoding and the lack of meaning and enabling parallel whole-form processing. Interestingly, utterances registered for novel ambiguous words indicate that these stimuli were likely represented as non-native words. Indeed, ambiguous novel words were mainly read in English rather in Russian language, a reading choice that did not change as a consequence of the training. Such preference for reading ambiguous novel words using L2 decoding rules could be tentatively explained by the learning context of the task, in which novel stimuli may tend to be considered as unknown L2 words rather than as unknown L1 (since the participant may assume familiarity with more words in L1 than in L2). Familiar words presented in the ambiguous script condition showed, in contrast, a different pattern of reading, since these stimuli were more often read as Russian words. This could be related to a more robust representation of these stimuli in readers' L1 lexicon, likely due to more exposure to them within native Russian rather than within English reading contexts. Nonetheless, familiar ambiguous words still exhibited a much less marked preference than novel ones for reading in one over another language. Indeed, analysis of the switching pattern between languages showed that familiar words exhibited a dynamic pattern of reading whereby words were interchangeably read either in Russian or English language across the exposures, suggesting strong orthographic decoding skills in both languages. Novel ambiguous word forms, however, showed a more static reading pattern with a smaller number of switches from one language to another, mainly read ─and hence represented─ in L2 orthographic lexicon.

Likely, training these stimuli in more ecological reading contexts (e.g., with novel ambiguous words embedded in L1 and L2 texts) would facilitate their progressive representation in both languages, resulting in their efficient and dynamic recognition across both L1 and L2, as observed for familiar ambiguous words.

Second, our post-training data also confirm efficient access, and thus acquisition of orthographic representations, for novel ambiguous words, as these stimuli showed similar performance for familiar words both in recall and recognition tasks, despite the associated grapheme-to-phoneme ambiguity. Moreover, ambiguous novel words were recognized faster than untrained ambiguous novel word forms presented as foils. This finding indicates the presence of a successful orthographic learning, similarly to previous studies using orthographic choice tasks in unambiguous L1 scripts (Share, 1999; Cunningham, Perry, Stanovich, & Share, 2002). Furthermore, our lexical decision task data also suggest the formation of stable orthographic representations of the newly-trained novel words regardless of phonological inconsistencies. The responses to these words took longer and were more error-prone than the rejection of novel but untrained words. This pattern, in agreement with that found in previous studies (Merkx, Rastle, & Davis, 2011; Leminen, Kimppa, Leminen, Lehtonen, Mäkelä & Shtyrov, 2016), indicates an interference during the categorization of previously trained-words as non-lexical items, thus suggesting some level of memory representation for these stimuli, not evident for untrained foils that lack representations in the readers' mental lexicon.

Overall, results from both training and post-training tasks demonstrate that, although phonological ambiguity interferes with the efficient decoding of novel words, such effect does not prevent the successful development of orthographic representation in reader's lexicon. These findings contradict one of our initial hypotheses predicting poorer achievement of orthographic learning for novel words with ambiguous L2/L1 graphemes. Nonetheless, the build-up of orthographic representations for ambiguous words must be examined carefully not only at the end, but also along the training. This online analysis showed that the process of building an orthographic representation was actually delayed for those words with inconsistent grapheme-to-phoneme decoding: whereas novel words presented in consistent L1 and L2 scripts matched the naming latencies for familiar words as early as at their second exposure (in case of those in L2) or third exposure (for those in L1), the naming of novel ambiguous words became fully comparable to the familiar words only at the tenth presentation. Such effect of phonological inconsistency in the orthographic learning of novel words may not be surprising considering the critical role of phonological decoding in the acquisition of reading and in novel word learning, as particularly evident when this skill is limited or impaired (Perfetti, 2003; Share & Stanovich, 1997; Kyte & Johnson, 2006). In essence, our prediction of poorer performance was not entirely incorrect, but could now be specified based on the result as a poorer acquisition *speed*, yet leading to the same ultimate performance. Importantly, these findings show, for the first time, the impact of phonological inconsistencies across L1 and L2 alphabets on the acquisition of new vocabulary during reading.

The successful achievement of orthographic learning of novel words under conditions of phonological inconsistency could be tentatively explained by the availability of two different reading strategies, each corresponding to the decoding principles of L1 and L2 alphabets. Encountering an orthographic ambiguity activates corresponding grapheme-to-phoneme decoding rules in both languages and then forces the reader to choose a decoding strategy (which, for novel items, resulted more often in the selection of L2 decoding rules); as a consequence, the processing of these stimuli becomes more effortful, increasing the attentional resources allocated to reading them, which is especially true in a single-word presentation paradigm, where there are no rules or prompts that could facilitate the election of a reading strategy. Although this leads to the naming latencies' inflation, it may also lead to a deeper memory encoding. Indeed, the script effect obtained in the recall task corroborates this interpretation, showing *better retrieval* for stimuli with ambiguous grapheme-to-phoneme decoding (both familiar and novel), in comparison to those presented in consistent L1 Cyrillic or L2 Roman graphemes, and thus suggesting the presence of stronger orthographic representations for the words whose reading is possible following not one, but two different decoding strategies. Therefore, although the knowledge of two alphabets may interfere with the phonological decoding of words with ambiguous phonemic correspondences across alphabets, it ultimately seems to contribute to their better memory and learning, particularly to the retention and access of their memory traces. Such effect, previously described as a *desirable difficulty effect* (Bjork, 1994), has been consistently found across different contexts including reading (Diemand-Yauman, Oppenheimer & Vaughan, 2011) and lexical acquisition (Eskenazi & Nix, 2020), showing better memory retention of stimuli as a consequence of their effortful encoding, by means of their presentation in hard-to-read fonts or, in this case, in phonologically ambiguous graphemes across L1-L2. Therefore, although the processing of words that can be read across different languages is effortful at encoding, such effort turns into beneficial or *desirable*, as it leads to advantageous retention enhancements. This view agrees with previous studies showing the role of biliteracy in reading and novel word learning, the so-called *biliteracy advantage*. In these studies, biliterates show better orthographic learning than bilingual monoliterate learners, likely as a consequence of a higher flexibility of their orthographic systems (Kahn-Horwitz et al., 2014; Modirkhamene, 2006; Schwartz, Geva, Share & Leikin, 2007; Schwartz et al., 2014). Notably, in these studies the facilitation was based on the transfer of orthographic characteristics across languages using non-overlapping writing systems in the absence of any decoding conflict. In contrast, in the present research the biliteracy advantage appears to be driven by inconsistencies across orthographic scripts, leading to higher control and monitoring of decoding processes and, consequently, stronger retention of novel words. This suggestion that inconsistencies across overlapping alphabets could lead to advantages in orthographic learning should still be further explored in future research, for instance by a direct comparison of mono vs. biliterate populations and of overlapping vs. non-overlapping L1/L2 combinations in the same individuals.

Regarding the process of orthographic learning under conditions of phonological consistency, results found in this study confirmed the effective acquisition of orthographic representations for novel words with non-ambiguous graphemes, both in the native and non-native alphabets. Indeed, naming latencies for novel words both in L1 Cyrillic and L2 Roman scripts decreased significantly across their training and matched those obtained for familiar words very quickly, already after two or three exposures, indicating similar reading automatization in native and non-native scripts. Moreover, in both alphabets, the performance for novel words in recall and recognition tasks did not differ from the one exhibited by familiar words, indicating efficient access to newly-formed memory traces after training. These results confirm previous findings in biliterate population, reflecting the rapid acquisition of new vocabulary through L2 reading (Chung, Chen, Commissaire, Krenca & Deacon, 2019; Schwartz, Kahn-Horwitz, & Share, 2014; van Daal & Wass, 2017). Importantly, the present study extends these findings by addressing this topic in bialphabetic population and comparing orthographic learning under native and non-native scripts, highlighting the role of phonological inconsistencies in this process. Nonetheless, more research is needed to further understand reading and orthographic learning processes in biliterate population, particularly in regards to the underlying brain mechanisms that support the successful representation of novel words even in conditions of phonological inconsistency. To this end, future studies could conduct a similar experimental design to the one carried out here and complement it with online recordings of brain activity during the training using, e.g., EEG or MEG paradigms that are known to be sensitive to orthographic memory trace activations (Bakker, Takashima, van Hell, Janzen, & McQueen, 2015; McLaughlin, Osterhout, & Kim, 2004; Mestres-Missé, Rodriguez-

Fornells, & Münte, 2007; Partanen, Leminen, Cook, & Shtyrov, 2018). Moreover, future investigation might consider manipulating the context of training, by repeatedly presenting novel words either in isolation or in association with a semantic reference (for instance, by associating them with pictures or photographs of novel object, or embedding them in sentences). Importantly, such manipulation would go beyond the study of purely orthographic learning addressed in the present work, and could investigate the acquisition of fully-fledge lexical representations through the integration of word's features at orthographic, phonological and semantic levels of information. Likely, meaningful training conditions would facilitate novel word learning, as has been suggested in previous studies (Angwin, Phua, & Copland, 2014; Bermúdez-Margaretto, Beltrán, Shtyrov, Dominguez & Cuetos, 2020; Havas, Taylor, Vaquero, de Diego-Balaguer, Rodríguez-Fornells, & Davis, 2018; Takashima, Bakker, Van Hell, Janzen, & McQueen, 2014), and would potentially reduce the impact of phonological inconsistencies led by graphemic overlap across L1-L2. Importantly, the number of novel words used in such studies should be kept to a minimum (as in the present experiment), thus ensuring both limited variability and a strict control across conditions as well as preventing memory overload in learners. While providing obvious methodological advantages (which is particularly important for new designs not used previously, such as the one here), this practice, may be associated with a diminished statistical power, which is usually mitigated by conducting a priori power analysis. In the present study we used observed power instead to ensure that our data had sufficient statistical power. At the same time, we are conscious of the fact that the observed power often overestimates the true power of a study (Hoenig & Heisey, 2001; Vasishth & Nicenboim, 2016, pp. 359); hence it must be considered as a limitation in the present study, which should be overcome in future research by conducting a priori power analysis.

The level of L2 proficiency was assessed in the present study by means of the LEAP-Q questionnaire, which is based on participant's self-ratings. This has been repeatedly demonstrated to be a sensitive tool for the assessment of L2 proficiency and experience across different languages. Importantly, the criterion-based validity of LEAP-Q was successfully established by confirming a relationship between self-reported data on this questionnaire and participants' performance on objective, behavioural measures (such as reading fluency, vocabulary or grammaticality judgment tasks) – both in L1 and L2 (Marian, Blumenfeld & Kaushanskaya, 2007). Nevertheless, some previous findings suggest that self-ratings might be less reliable than objective L2 proficiency measures (e.g., Khare, Verma, Kar, Srinivasan, & Brysbaert, 2013; Lemhöfer & Broersma, 2012; Wen & van Heuven, 2017). Indeed, many existing studies (e.g., FitzPatrick & Indefrey, 2010; Lemhöfer & Broersma, 2012; van der Meij, Cuetos, Carreiras & Barber, 2011) have used more objective L2 proficiency measures in addition to participants' self-ratings, which is especially important if these metrics are used to predict experimental data. Therefore, the sole use of LEAP-Q in our work may still be considered as a limitation. Future studies on word learning (as well as on language processing in general) which involve biliterate or bilingual populations should combine validated and standardized self-assessment questionnaires with objective measures aiming to better predict the impact of L2 proficiency on experimental data (e.g., Lemhöfer & Broersma, 2012).

Regarding the stimuli trained in the native alphabet, these exhibited generally better reading performance than those presented in the non-native script. That is, novel words in L2 alphabet showed slower naming latencies than those in L1 script despite the training. This second alphabet effect is not surprising considering that our sample consisted of unbalanced Russian-English biliterates who learnt English and the corresponding Roman alphabet during their late childhood, as reflected by the high age of acquisition. Consequently, language proficiency and exposure to scripts were also unbalanced, giving us an additional leverage in comparing L1 and L2 orthographic learning performance. A different pattern of results could be expected in a better-balanced group of early biliterates where it would be reasonable to expect not only a smaller L2 alphabet effect but also a reduced phonological inconsistency leading to a more efficient L2 word learning. Indeed, results of regression analyses partially support this view (see supplementary material section) as L2 proficiency and exposure predicted faster decoding after training — both under consistent and inconsistent script conditions. Future research will need to compare orthographic learning skills across early and late bialphabetic learners to corroborate this preliminary finding.

Finally, it is worthwhile to highlight that the present research provides empirical evidence about reading and orthographic learning skills in Russian (as particularly exhibited by stimuli presented in non-ambiguous, L1 Cyrillic script). Since this language is highly similar to many other alphabetic languages in terms of orthographic representation and depth, and is closely related to many Indo-European languages, results reported here may contribute not only to the investigation of Russian language but, importantly, could be generalizable to other languages as well. Indeed, the majority of studies addressing orthographic learning and related topics are conducted in handful of Western European languages, predominantly English, despite the fact that this language has several characteristics that make it an outlier in comparison to many other languages across the world (most importantly, its extreme orthographic depth with poor transparency between written and spoken forms) which question the conventional generalization of English-based results to other languages (Seymour, Aro & Squirne, 2009; Share, 2008; Ziegler & Goswami, 2005). Thus, the present study extends this research, providing evidence for decoding and novel word learning process in a rather transparent orthography, comparable to many other languages and scripts within the world's family of alphabetic systems.

Declarations

Funding

The study was funded by a grant from the Russian Science Foundation (project No. 19-78-00140) awarded to HSE University.

Conflicts of interest/Competing interests

The authors declare no conflict of interest.

Ethics approval

The manuscript does not contain clinical studies or patient data. The study was approved by the Ethics Committee of the Department of Psychology, National Research University Higher School of Economics.

Consent to participate (include appropriate statements)

All participants gave their written consent to take part in the study.

Consent for publication

All authors have approved the manuscript and agree with its publication.

Availability of data and material

The data that support the findings of this study are available from the corresponding author upon reasonable request.

Code availability

The code used in this study is available from the corresponding author upon reasonable request.

References

Abdelhadi, S., Ibrahim, R., & Eviatar, Z. (2011). Perceptual load in the reading of Arabic: Effects of orthographic visual complexity on detection. *Writing Systems Research*, *3*(2), 117–127.

Álvarez-Cañizo, M., Suárez-Coalla, P., & Cuetos, F. (2019). Orthographic learning in Spanish children: influence of previous semantic and phonological knowledge. *Journal of Research in Reading*, *42*(1), 137–149. https://doi.org/10.1111/1467-9817.12254

ÁLVAREZ-CAÑIZO, M., SUÁREZ-COALLA, P., & CUETOS, F. (2018). The role of sublexical variables in reading fluency development among Spanish children. *Journal of Child Language*, *45*(4), 858–877. https://doi.org/10.1017/S0305000917000514

Ando, E., Jared, D., Nakayama, M., & Hino, Y. (2014). Cross-script phonological priming with Japanese Kanji primes and English targets. *Journal of Cognitive Psychology*, 26(8), 853-870.

Angwin, A. J., Phua, B., & Copland, D. A. (2014). Using semantics to enhance new word learning: An ERP investigation. Neuropsychologia, 59, 169-178.

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, *59*(4), 390–412.

Bakker, I., Takashima, A., van Hell, J. G., Janzen, G., & McQueen, J. M. (2015). Tracking lexical consolidation with ERPs: Lexical and semantic-priming effects on N400 and LPC responses to newly-learned words. Neuropsychologia, 79, 33-41.

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

Bates, D., Kliegl, R., Vasishth, S., & Baayen, H. (2015). Parsimonious mixed models. arXiv preprint arXiv:1506.04967.

Bermúdez-Margaretto, B., Beltrán, D., Shtyrov, Y., Dominguez, A., & Cuetos, F. (2020). Neurophysiological Correlates of Top-Down Phonological and Semantic Influence during the Orthographic Processing of Novel Visual Word-Forms. Brain sciences, 10(10), 717.

Bjork, R. A. (1994). Memory and metamemory considerations in the. Metacognition: Knowing about knowing, 185.

Boersma, P. (2011). Praat: doing phonetics by computer [Computer program]. *Http://Www. Praat. Org/*.

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.

Brysbaert, M., Van Dyck, G., & Van de Poel, M. (1999). Visual word recognition in bilinguals: evidence from masked phonological priming. *Journal of Experimental Psychology: Human Perception and Performance*, 25(1), 137.

Bultena, S., Dijkstra, T., & van Hell, J. G. (2013). Cognate and word class ambiguity effects in noun and verb processing. *Language and Cognitive Processes*, *28*(9), 1350–1377.

Chung, S. C., Chen, X., & Geva, E. (2019). Deconstructing and reconstructing cross-language transfer in bilingual reading development: An interactive framework. *Journal of Neurolinguistics*, *50*, 149–161.

Clay, F., Bowers, J. S., Davis, C. J., & Hanley, D. A. (2007). Teaching adults new words:

The role of practice and consolidation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *33*(5), 970.

Cop, U., Dirix, N., Van Assche, E., Drieghe, D., & Duyck, W. (2017). Reading a book in one or two languages? An eye movement study of cognate facilitation in L1 and L2 reading. *Bilingualism: Language and Cognition*, *20*(4), 747–769.

Cunningham, A. E., Perry, K. E., Stanovich, K. E., & Share, D. L. (2002). Orthographic learning during reading: Examining the role of self-teaching. *Journal of Experimental Child Psychology*, *82*(3), 185–199. https://doi.org/10.1016/S0022-0965(02)00008-5

de Jong, P. F., & Share, D. L. (2007). Orthographic learning during oral and silent reading. *Scientific Studies of Reading*, *11*(1), 55–71.

Diemand-Yauman, C., Oppenheimer, D. M., & Vaughan, E. B. (2011). Fortune favors the (): Effects of disfluency on educational outcomes. *Cognition*, 118(1), 111-115.

Ehri, L. C. (1992). Reconceptualizing the development of sight word reading and its relationship to recoding.

Eskenazi, M. A., & Nix, B. (2020). Individual differences in the desirable difficulty effect during lexical acquisition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*.

FitzPatrick, I., & Indefrey, P. (2010). Lexical competition in nonnative speech comprehension. Journal of Cognitive Neuroscience, 22(6), 1165-1178.

Forster, K. I., & Chambers, S. M. (1973). Lexical access and naming time. *Journal of Verbal Learning and Verbal Behavior*, *12*(6), 627–635.

Fox, J., Weisberg, S., Friendly, M., Hong, J., Andersen, R., Firth, D., ... & Fox, M. J. (2016). Package ‘effects’. Url: http://www. r-project. org, http://socserv. socsci. mcmaster.

ca/jfox.

Goswami, U., & Bryant, P. (2016). Phonological skills and learning to read. Psychology Press.

Havas, V., Taylor, J. S. H., Vaquero, L., de Diego-Balaguer, R., Rodríguez-Fornells, A., & Davis, M. H. (2018). Semantic and phonological schema influence spoken word learning and overnight consolidation. Quarterly Journal of Experimental Psychology, 71(6), 1469-1481.

Hamada, M., & Koda, K. (2008). Influence of first language orthographic experience on second language decoding and word learning. *Language Learning*, *58*(1), 1–31.

Havelka, J., & Rastle, K. (2005). The assembly of phonology from print is serial and subject to strategic control: evidence from serbian. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *31*(1), 148.

Hoenig, J. M., and D. M. Heisey. 2001. The abuse of power: the pervasive fallacy of power calculations for data analysis. The American Statistician 55.1 19–24

Jared, D., & Kroll, J. F. (2001). Do bilinguals activate phonological representations in one or both of their languages when naming words?. *Journal of memory and language*, 44(1), 2-31.

Jared, D., & Szucs, C. (2002). Phonological activation in bilinguals: Evidence from interlingual homograph naming. *Bilingualism*, 5(3), 225.

Jeffreys, H. (1961). Theory of probability (3rd ed.). Oxford: Oxford University Press

Jouravlev, O., Lupker, S. J., & Jared, D. (2014). Cross-language phonological activation: Evidence from masked onset priming and ERPs. *Brain and language*, 134, 11-22.

Kahn-Horwitz, J., Kuash, S., Ibrahim, R., & Schwartz, M. (2014). How do previously acquired languages affect acquisition of English as a foreign language: The case of Circassian. *Written Language & Literacy*, *17*(1), 40–61.

Khare, V., Verma, A., Kar, B., Srinivasan, N., & Brysbaert, M. (2013). Bilingualism and the increased attentional blink effect: Evidence that the difference between bilinguals and monolinguals generalizes to different levels of second language proficiency. Psychological Research, 77(6), 728-737.

Kim, J., & Davis, C. (2003). Task effects in masked cross-script translation and phonological priming. *Journal of Memory and Language*, 49(4), 484-499.

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest package: tests in linear mixed effects models. *Journal of Statistical Software*, *82*(13).

Kwok, R. K. W., Cuetos, F., Avdyli, R., & Ellis, A. W. (2017). Reading and lexicalization in opaque and transparent orthographies: Word naming and word learning in English and Spanish. *The Quarterly Journal of Experimental Psychology*, *70*(10), 2105–2129. https://doi.org/10.1080/17470218.2016.1223705

Kwok, R. K. W., & Ellis, A. W. (2015). Visual word learning in skilled readers of English. *The Quarterly Journal of Experimental Psychology*, *68*(2), 326–349. https://doi.org/10.1080/17470218.2014.944549

Kyte, C. S., & Johnson, C. J. (2006). The role of phonological recoding in orthographic learning. *Journal of Experimental Child Psychology*, *93*(2), 166–185.

Lallier, M., & Carreiras, M. (2018). Cross-linguistic transfer in bilinguals reading in two alphabetic orthographies: The grain size accommodation hypothesis. *Psychonomic Bulletin & Review*, *25*(1), 386–401.

Lemhöfer, K., & Broersma, M. (2012). Introducing LexTALE: A quick and valid lexical test for advanced learners of English. Behavior research methods, 44(2), 325-343.

Leminen, A., Kimppa, L., Leminen, M. M., Lehtonen, M., Mäkelä, J. P., & Shtyrov, Y. (2016). Acquisition and consolidation of novel morphology in human neocortex: A neuromagnetic study. Cortex, 83, 1-16.

Lukatela, G. (1999). Effects of frequency and phonological ambiguity on naming Serbo-Croatian words. *European Journal of Cognitive Psychology*, *11*(1), 1–16.

Maloney, E., Risko, E. F., O'Malley, S., & Besner, D. (2009). Short Article: Tracking the Transition from Sublexical to Lexical Processing: On the Creation of Orthographic and Phonological Lexical Representations. *Quarterly Journal of Experimental Psychology*, *62*(5), 858–867. https://doi.org/10.1080/17470210802578385

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*.

Martens, V. E. G., & De Jong, P. F. (2008). Effects of repeated reading on the length effect in word and pseudoword reading. *Journal of Research in Reading*, *31*(1), 40–54.

McBride-Chang, C., Zhou, Y., Cho, J.-R., Aram, D., Levin, I., & Tolchinsky, L. (2011). Visual spatial skill: A consequence of learning to read? *Journal of Experimental Child Psychology*, *109*(2), 256–262.

McLaughlin, J., Osterhout, L., & Kim, A. (2004). Neural correlates of second-language word learning: Minimal instruction produces rapid change. Nature neuroscience, 7(7), 703-704.

Merkx, M., Rastle, K., & Davis, M. H. (2011). The acquisition of morphological knowledge investigated through artificial language learning. Quarterly Journal of Experimental Psychology, 64(6), 1200-1220.

Mestres-Missé, A., Rodriguez-Fornells, A., & Münte, T. F. (2007). Watching the Brain during Meaning Acquisition. Cerebral Cortex, 17(8), 1858-1866. Retrieved from http://cercor.oxfordjournals.org/content/17/8/1858.abstract.

Modirkhamene, S. (2006). The reading achievement of third language versus second language learners of English in relation to the interdependence hypothesis. *International Journal of Multilingualism*, *3*(4), 280–295.

Nag, S. (2007). Early reading in Kannada: The pace of acquisition of orthographic knowledge and phonemic awareness. *Journal of Research in Reading*, *30*(1), 7–22.

Nakayama, M., Sears, C. R., Hino, Y., & Lupker, S. J. (2012). Cross-script phonological priming for Japanese-English bilinguals: Evidence for integrated phonological representations. *Language and Cognitive Processes*, 27(10), 1563-1583.

Partanen, E. J., Leminen, A., Cook, C., & Shtyrov, Y. (2018). Formation of neocortical memory circuits for unattended written word forms: neuromagnetic evidence. Scientific reports, 8(1), 15829.

Paulesu, E., McCrory, E., Fazio, F., Menoncello, L., Brunswick, N., Cappa, S. F., Cotelli, M., Cossu, G., Corte, F., & Lorusso, M. (2000). A cultural effect on brain function. *Nature Neuroscience*, *3*(1), 91–96.

Peeters, D., Dijkstra, T., & Grainger, J. (2013). The representation and processing of identical cognates by late bilinguals: RT and ERP effects. *Journal of Memory and Language*, *68*(4), 315–332.

Perfetti, C. A. (2003). The Universal Grammar of Reading. *Scientific Studies of Reading*, *7*(1), 3–24. https://doi.org/10.1207/S1532799XSSR0701_02

Perfetti, C. A., & Dunlap, S. (2008). Learning to read: General principles and writing system variations. In *Learning to read across languages* (pp. 25–50). Routledge.

Perfetti, C. A., Liu, Y., & Tan, L.-H. (2002). How the mind can meet the brain in reading: A comparative writing systems approach. *Cognitive Neuroscience Studies of the Chinese Language*, 35–60.

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

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

Rastle, K., Havelka, J., Wydell, T. N., Coltheart, M., & Besner, D. (2009). The cross-script length effect: Further evidence challenging PDP models of reading aloud. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *35*(1), 238.

Reverberi, C., Kuhlen, A. K., Seyed-Allaei, S., Greulich, R. S., Costa, A., Abutalebi, J., & Haynes, J. D. (2018). The neural basis of free language choice in bilingual speakers: Disentangling language choice and language execution. NeuroImage, 177, 108-116. Salasoo, A., Shiffrin, R. M., & Feustel, T. C. (1985). Building permanent memory codes: codification and repetition effects in word identification. *Journal of Experimental Psychology: General*, *114*(1), 50.

Schad, D. J., Vasishth, S., Hohenstein, S., & Kliegl, R. (2020). How to capitalize on a priori contrasts in linear (mixed) models: A tutorial. Journal of Memory and Language, 110, 104038.

Schneider, W., Eschman, A., & Zuccolotto, A. (2002). *E-Prime: User's guide*. Psychology

Software Incorporated.

Schwartz, M., Geva, E., Share, D. L., & Leikin, M. (2007). Learning to read in English as third language: The cross-linguistic transfer of phonological processing skills. *Written Language & Literacy*, *10*(1), 25–52.

Schwartz, M., Kahn-Horwitz, J., & Share, D. L. (2014). Orthographic learning and self-teaching in a bilingual and biliterate context. *Journal of Experimental Child Psychology*, *117*(1), 45–58. https://doi.org/10.1016/j.jecp.2013.08.008

Seymour, P. H. K., Aro, M., Erskine, J. M., & Network, C. with C. A. A. (2003). Foundation literacy acquisition in European orthographies. *British Journal of Psychology*, *94*(2), 143–174.

Share, D L, & Stanovich, K. E. (1995). *Cognitive processes in early reading development: accommodating activation associated with word recognition in children with reading difficulties*.

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

Share, David L. (2008). On the Anglocentricities of current reading research and practice: The perils of overreliance on an" outlier" orthography. *Psychological Bulletin*, *134*(4), 584.

Share, D. L. (2008). Orthographic learning, phonological recoding, and self-teaching. In Advances in child development and behavior (Vol. 36, pp. 31-82). JAI.

Snowling, M. J., & Göbel, S. M. (2011). Reading development and dyslexia.

Suárez-Coalla, P., Álvarez-Cañizo, M., & Cuetos, F. (2016). Orthographic learning in Spanish children. *Journal of Research in Reading*, *39*(3), 292–311.

Takashima, A., Bakker, I., Van Hell, J. G., Janzen, G., & McQueen, J. M. (2014). Richness of information about novel words influences how episodic and semantic memory networks interact during lexicalization. NeuroImage, 84, 265-278.

Tamura, N., Castles, A., & Nation, K. (2017). Orthographic learning, fast and slow: Lexical competition effects reveal the time course of word learning in developing readers. *Cognition*, *163*, 93–102.

Team, R. C. (2013). *R: A language and environment for statistical computing*.

van Daal, V. H. P., & Wass, M. (2017). First- and Second-Language Learnability Explained by Orthographic Depth and Orthographic Learning: A “Natural” Scandinavian Experiment. *Scientific Studies of Reading*, *21*(1), 46–59.

Van Der Meij, M., Cuetos, F., Carreiras, M., & Barber, H. A. (2011). Electrophysiological correlates of language switching in second language learners. Psychophysiology, 48(1), 44-54.

Vasishth, S., & Nicenboim, B. (2016). Statistical methods for linguistic research: Foundational ideas–Part I. Language and Linguistics Compass, 10(8), 349-369.

Von Studnitz, R. E., & Green, D. W. (2002). Interlingual homograph interference in German-English bilinguals: Its modulation and locus of control. *Bilingualism*, 5(1), 1.

Wagenmakers, E. J. (2007). Stopping rules and their irrelevance for Bayesian inference: Online appendix to “A practical solution to the pervasive problems of p–values.”. Psychonomic Bulletin and Review, 1-5.

Wang, H.-C., Castles, A., & Nickels, L. (2012). *Word regularity affects orthographic learning*. SAGE Publications Sage UK: London, England.

Wen, Y., & van Heuven, W. J. (2017). Non-cognate translation priming in masked priming lexical decision experiments: A meta-analysis. Psychonomic Bulletin & Review, 24(3), 879-886.

Zhou, H., Chen, B., Yang, M., & Dunlap, S. (2010). Language nonselective access to phonological representations: Evidence from Chinese–English bilinguals. Quarterly Journal of Experimental Psychology, 63(10), 2051-2066.

Ziegler, J. C., & Goswami, U. (2005). Reading acquisition, developmental dyslexia, and skilled reading across languages: a psycholinguistic grain size theory. *Psychological Bulletin*, *131*(1), 3.

Table 1. Participants' second language (English) proficiency evaluation obtained by means of LEAP-Q Questionnaire.

	Mean	**SD, range**
General L2 proficiency (0-100)	66.67	20.14, 83.33
L2 proficiency (0-10)		
Listening	6.92	2.00, 8
Speaking	6.16	2.23, 9
Reading	6.92	2.00, 8
General L2 Exposure (0-100)	44.67	15.48, 66.67
L2 Exposure (0-10)		
Friends	6.16	2.23, 9
Family	7.42	1.89, 8
Reading	2.96	2.16, 9
Language tapes / Self-instruction	0.6	0.98, 3
TV / Media	6.36	2.54, 10
Radio / Music	3.78	3.07, 9
Factors contributing to L2 learning (0-10)		
Friends	3.78	2.87, 10
Family	1.72	2.43, 8
Reading	6.9	2.50, 10
Language tapes / Self-instruction	4.64	3.21, 10
TV / Media	4.32	3.12, 10
Radio / Music	6.4	2.49, 9
Years immerse in L2 environment		
Country	0.58	1.40, 7
Family	0.09	0.31, 1.83
School/work	0.92	2.62, 12
Age of L2 acquisition	7.82	2.71, 13
Age of L2 fluency onset	14.77	3.15, 15
Age of L2 reading acquisition	11.59	3.64, 15
Age of L2 reading fluency onset	15.09	2.90, 12
Number of non-native languages learnt	3.18	0.99, 3

Table 2. Mean and standard error of naming latencies (RTs) for familiar and novel word forms in different scripts at the first and the last block of the training task.

	Block 1		**Block 10**	
	Familiar	**Novel**	**Familiar**	**Novel**
L1	632 (26.3)	761 (26.4)	552 (26.4)	575 (26.3)
L2	809 (26.4)	936 (26.4)	628 (26.3)	651 (26.4)
Ambiguous	789 (26.3)	1068 (29.7)	666 (26.3)	735 (29.6)

Table 3. Statistical results obtained for response latencies collected during the training (reading aloud task).

Global analysis (latency in reading aloud task)

Block	*F*(9,10947.5)=141.47, *p*<.001
Lexicality	*F*(1, 22.2)=18.32, *p*<.001
Script	*F*(2,25.1)=29.20, *p*<.001
Script x Lexicality	*F*(2,17)=3.30, *p*=.061
Block x Lexicality	*F*(9,10947.5)=18.80, *p*<.001
Block x Script	*F*(18,10947.4)=5.49, *p*<.001
Block x Lexicality x Script	*F*(18,10947.3)=1.89, *p*=.012

Random effects

Groups name	Variance	Std. Dev.	Corr.
Participant (Intercept)	9577.2	97.86	
Lexicality	500.9	22.38	0.58
Script 1 (English)	2618.4	51.17	0.48 0.39
Script 2 (Russian)	876.0	29.60	0.09 0.14 -0.57
Item (Intercept)	1473.1	38.38	
Residual	26892.3	163.99	

Number of obs.: 11219; groups: participant, 50; items, 23

Follow-up comparisons

Script x Lexicality across blocks

	Block 1				Block 10			
	Estimate	SE	z	p	Estimate	SE	z	p
L2 vs. L1 in Familiar	177.2	32.6	5.436	<.0001	75.7	32.6	2.320	0.0407
Ambiguous vs. L1 in Familiar	156.9	34.1	4.597	<.0001	115.2	34.2	3.370	0.0023
Ambiguous vs. L2 in Familiar	-20.3	33.4	0.608	0.5434	39.5	33.4	1.182	0.2370
L2 vs. L1 in Novel	176.1	32.7	5.393	<.0001	75.8	32.6	2.326	0.0200
Ambiguous vs. L1 in Novel	307.1	36.8	8.345	<.0001	159.8	36.7	4.350	<.0001
Ambiguous vs. L2 in Novel	131.0	36.1	3.631	0.0003	83.9	36.0	2.331	0.0200
Novel vs Familiar in L1	129.3	32.4	3.993	0.0001	23.5	32.4	0.724	0.4688
Novel vs. Familiar in L2	128.2	32.5	3.949	0.0001	23.6	32.4	0.728	0.4666
Novel vs. Familiar in Ambiguous	279.5	35.1	7.953	<.0001	68.1	35.1	1.940	0.0524

Lexicality x Block across scripts

	L1				L2				Ambiguous			
	Estimate	SE	z	p	Estimate	SE	z	p	Estimate	SE	z	p
Block 1 vs. Block 10 in Familiar	80.45	16.5	4.881	<.0001	181.96	16.6	10.969	<.0001	122.21	16.5	7.415	<.0001
Block 1 vs. Block 10 in Novel	186.26	16.5	11.272	<.0001	286.53	16.6	17.248	<.0001	333.62	19.9	16.794	<.0001
Novel vs. Familiar in Block 1	129.3	32.4	3.993	0.0001	128.2	32.5	3.949	0.0001	279.5	35.1	7.953	<.0001
Novel vs. Familiar in Block 10	23.5	32.4	0.724	0.4688	23.6	32.4	0.728	0.4666	68.1	35.1	1.940	0.0524

Block x Scripts across Lexicality

	Familiar				Novel			
	Estimate	SE	z	p	Estimate	SE	z	p
Block 1 vs. Block 10 in L1	80.45	16.5	4.881	<.0001	186.26	16.5	11.272	<.0001
Block 1 vs. Block 10 in L2	181.96	16.6	10.969	<.0001	286.53	16.6	17.248	<.0001
Block 1 vs. Block 10 in Ambiguous	122.21	16.5	7.415	<.0001	333.62	19.9	16.794	<.0001
L2 vs. L1 in Block 1	177.2	32.6	5.436	<.0001	176.1	32.7	5.393	<.0001
Ambiguous vs. L1 in Block 1	156.9	34.1	4.597	<.0001	307.1	36.8	8.345	<.0001
Ambiguous vs. L2 in Block 1	-20.3	33.4	-0.608	0.5434	131.0	36.1	3.631	0.0003
L2 vs. L1 in Block 10	75.7	32.6	2.320	0.0407	75.8	32.6	2.326	0.0200
Ambiguous vs. L1 in Block 10	115.2	34.2	3.370	0.0023	159.8	36.7	4.350	<.0001
Ambiguous vs. L2 in Block 10	39.5	33.4	1.182	0.2370	83.9	36.0	2.331	0.0200

Table 4. Lexicality effect obtained for each script condition and block across the training task.

	L1		**L2**		**Ambiguous**	
	Estimate	***p* value**	**Estimate**	***p* value**	**Estimate**	***p* value**
Block 1	129.3	.0001	128.2	.0001	279.50	<.0001***
Block 2	63.9	.0485*	63.0	.051	192.7	<.0001***
Block 3	31.2	.335	54.8	.091	187.4	<.0001***
Block 4	42.8	.186	29.3	.36	152.0	<.0001**
Block 5	43.9	.175	48.3	.13	105.5	.0027**
Block 6	31.1	.337	48.8	.13	121.8	.0005**
Block 7	14.3	.657	12.70	.70	96.5	.006**
Block 8	31.3	.333	34.7	.28	68.9	.0496*
Block 9	13.9	.668	25.5	.43	91.7	.0087**
Block 10	23.5	.468	23.63	.46	68.1	.052

***<.0001
**<.01
*<.05

Table 5. Statistical results obtained for response latencies collected in the recognition task.

Global analysis (latency in recognition task)

Lexicality	$F(1, 44.241)=12.44$, $p<.001$
Script	$F(2, 43.653)=0.0432$, $p>.05$
Familiarity	$F(1, 43.667)=11.62$, $p<.001$
Lexicality x Script	$F(2, 43.652)=0.68$, $p>.05$
Lexicality x Familiarity	$F(1, 43.639)=3.51$, $p=.067$
Script x Familiarity	$F(2, 43.657)=3.66$, $p<.05$
Lexicality x Script x Familiarity	$F(2, 43.657)=0.52$, $p>.05$

Random effects

Groups name	Variance	Std. Dev.	Corr.
Participant (Intercept)	7006.6	83.71	
Lexicality	188.3	13.72	-0.03
Item (Intercept)	4285.7	65.47	
Lexicality	5175.4	71.94	0.39
Residual	75389.8	274.57	

Number of obs.: 2864; groups: participant, 50; items, 69

Follow-up comparisons

	Lexicality x Familiarity				**Script x Familiarity**				
	Estimate	**SE**	**z**	**p**		**Estimate**	**SE**	**z**	**p**
Foils vs. Trained in Familiar	**42.3**	**30.8**	**1.370**	**0.1706**	Foils vs. Trained in L1	**15.3**	**46.3**	**0.331**	**0.7405**
Foils vs. Trained in Novel	**145.6**	**45.7**	**3.189**	**0.0014**	Foils vs. Trained in L2	**189.7**	**46.0**	**4.128**	**<.0001**
Novel vs. Familiar in Foils	**149.8**	**32.2**	**4.646**	**<.0001**	Foils vs. Trained in Amb.	**76.8**	**50.8**	**1.513**	**0.1304**
Novel vs. Familiar in Trained	**46.5**	**45.0**	**1.032**	**0.3021**	L2 vs. L1 in Foils	**86.9**	**38.0**	**2.289**	**0.0574**
					Ambiguous vs. L1 in Foils	**39.4**	**39.9**	**0.989**	**0.5836**
					Ambiguous vs. L2 in Foils	**-47.4**	**39.8**	**-1.193**	**0.4572**
					L2 vs. L1 in Trained	**-87.5**	**53.0**	**-1.649**	**0.2250**
					Ambiguous vs. L1 in Trained	**-22.0**	**56.0**	**-0.393**	**0.9182**
					Ambiguous vs. L2 in Trained	**65.5**	**55.7**	**1.174**	**0.4685**

Table 6. Statistical results obtained for response latencies collected in the lexical decision task.

Global analysis (latency in lexical decision task)

Lexicality	*F*(1, 70.592)=110.50, *p*<.001
Script	*F*(2, 49.779)=28.89, *p*<.001
Familiarity	*F*(1, 49.950)= 0.09, *p*>.05
Lexicality x Script	*F*(2, 49.782)= 6.15, *p*<.01
Lexicality x Familiarity	*F*(1, 49.961)= 12.97, *p*<.001
Script x Familiarity	*F*(2, 49.627)= 1.47, *p*>.05
Lexicality x Script x Familiarity	*F*(2, 49.632)= 2.59, *p*=.08

Random effects

Groups name	Variance	Std. Dev.	Corr.
Participant (Intercept)	13202	114.90	
Lexicality	5013	70.80	0.70
Item (Intercept)	1250	35.36	
Lexicality	5211	72.19	0.36
Residual	113405	336.76	

Number of obs.: 2573; groups: participant, 50; items, 69

Follow-up comparisons

	Lexicality x Familiarity				**Lexicality x Script**				
	Estimate	**SE**	**z**	**p**		**Estimate**	**SE**	**z**	**p**
Foils vs. Trained in Familiar	**100.1**	**31.0**	**3.230**	**0.0032**	Novel vs. Familiar in L1	**305**	**45.9**	**6.632**	**<.0001**
Foils vs. Trained in Novel	**-84.1**	**40.7**	**2.066**	**0.0479**	Novel vs. Familiar in L2	**251**	**48.0**	**5.237**	**<.0001**
Novel vs. Familiar in Foils	**250**	**36.0**	**6.943**	**<.0001**	Novel vs. Familiar in Amb.	**471**	**52.0**	**9.058**	**<.0001**
Novel vs. Familiar in Trained	**435**	**46.2**	**9.407**	**<.0001**	L2 vs. L1 in Familiar	**222**	**38.3**	**5.800**	**<.0001**
					Ambiguous vs. L1 in Familiar	**120**	**37.1**	**3.240**	**0.0088**
					Ambiguous vs. L2 in Familiar	**-102**	**38.6**	**-2.645**	**0.0339**
					L2 vs. L1 in Novel	**169**	**46.2**	**3.653**	**0.0032**
					Ambiguous vs. L1 in Novel	**287**	**51.3**	**5.586**	**<.0001**
					Ambiguous vs. L2 in Novel	**118**	**52.0**	**2.265**	**0.0761**

Table 7. Statistical results obtained for correct responses collected in the lexical decision task.

Global analysis (accuracy in lexical decision task)

Lexicality	$\chi 2$(1)= 0.70, *p*>.05
Script	$\chi 2$(2)= 20.55, *p*<.001
Familiarity	$\chi 2$(1)= 1.76, *p*>.05
Lexicality x Script	$\chi 2$(2)= 7.74, *p*<.05
Lexicality x Familiarity	$\chi 2$(1)= 21.48, *p*<.001
Script x Familiarity	$\chi 2$(2)= 1.68, *p*>.05
Lexicality x Script x Familiarity	$\chi 2$(2)= 3.11, *p*>.05

Random effects

Groups name	Variance	Std. Dev.	Corr.
Participant (Intercept)	0.5718	0.7562	
Lexicality	0.6300	0.7937	0.25
Item (Intercept)	0.6996	0.8364	
Lexicality	0.6350	0.7969	-0.46

Number of obs.: 2879; groups: participant, 50; items, 69

Follow-up comparisons

	Lexicality x Familiarity				**Lexicality x Script**				
	Estimate	**SE**	**z**	**p**		**Estimate**	**SE**	**z**	**p**
Foils vs. Trained in Familiar	**-2.35**	**0.663**	**-3.540**	**0.0004**	Novel vs. Familiar in L1	**-0.528**	**0.810**	**-0.651**	**0.5147**
Foils vs. Trained in Novel	**1.30**	**0.425**	**3.068**	**0.0022**	Novel vs. Familiar in L2	**0.895**	**0.650**	**1.378**	**0.1683**
Novel vs. Familiar in Foils	**1.41**	**0.522**	**2.698**	**0.0070**	Novel vs. Familiar in Amb.	**-1.620**	**0.761**	**-2.130**	**0.0332**
Novel vs. Familiar in Trained	**-2.24**	**0.728**	**-3.081**	**0.0021**	L1 vs. L2 in Familiar	**2.833**	**0.792**	**3.578**	**0.0010**
					Ambiguous vs. L1 in Familiar	**-0.998**	**0.861**	**-1.159**	**0.4777**
					Ambiguous vs. L2 in Familiar	**1.835**	**0.772**	**2.378**	**0.0458**
					L1 vs. L2 in Novel	**1.410**	**0.514**	**2.742**	**0.0168**
					Ambiguous vs. L1 in Novel	**-2.090**	**0.539**	**-3.880**	**0.0003**
					Ambiguous vs. L2 in Novel	**-0.680**	**0.491**	**-1.384**	**0.3491**

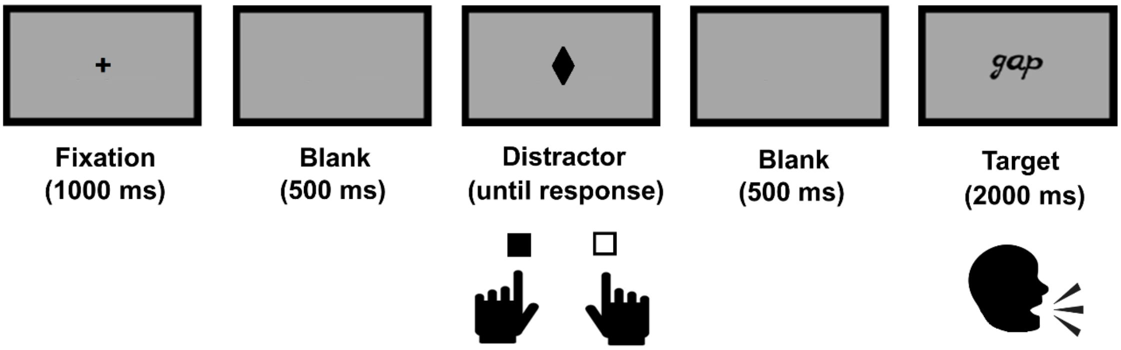

Figure 1. Sequence of stimuli presentation during the reading-aloud task (training phase).

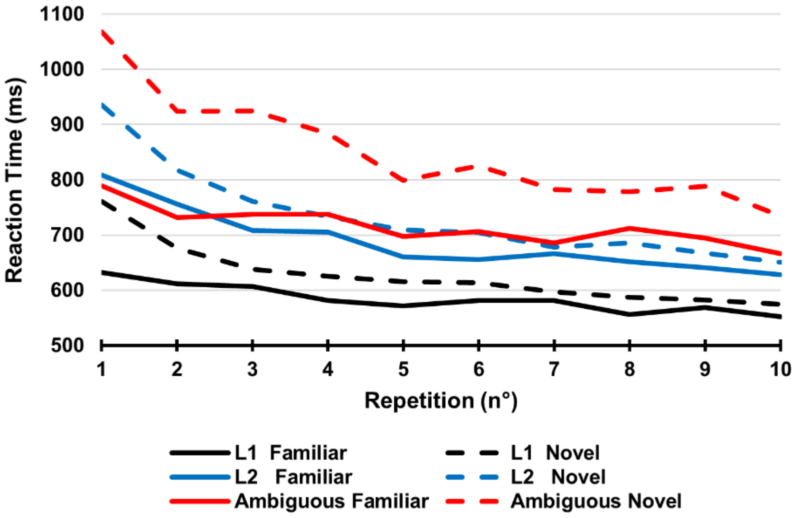

Figure 2. Mean naming latencies (RTs) obtained across training blocks for each experimental condition (familiar and novel words in L1, L2 and ambiguous scripts).

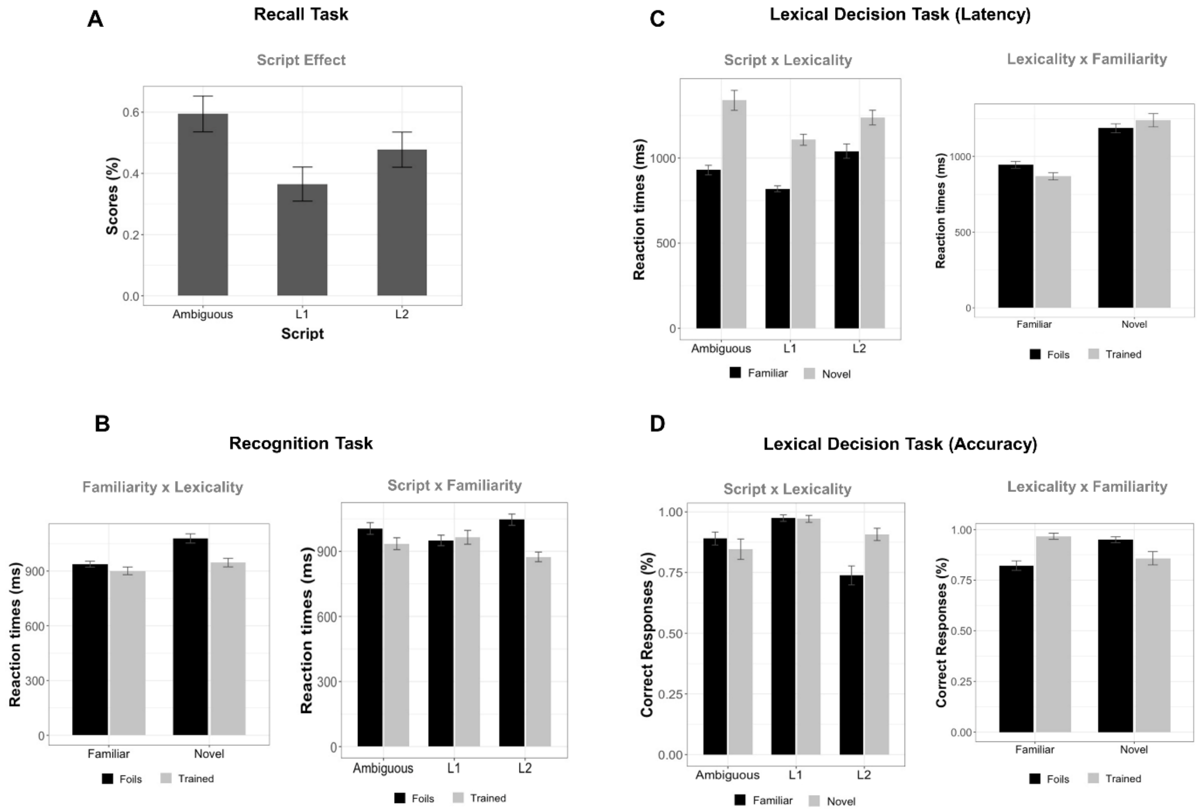

Figure 3. Results across the three post-training assessment tasks: free recall accuracy (A), recognition RT (B) and lexical decision task RT (C) and accuracy (D).

Appendix

1. Stimuli used in the study. Note that the same handwritten font was used across all tasks.

L1 Script		**L2 Script**		**Ambiguous Script**	
Familiar Words	**Novel Words**	**Familiar Words**	**Novel Words**	**Familiar Words**	**Novel Words**
шаг	шаз	kid	kof	дар	дех
шок	шой	law	leq	рот	рех
лак	лец	vet	vaz	сор	сик
зал	зеж	jam	jod	пар	пем