

**Investigating pseudohomophone interference effects in young second language (L2)
learners**

Running head: pseudohomophone effects in L2

Commissaire Eva, Duncan Lynne, Casalis Séverine

To appear in Journal of Experimental Child Psychology

Learning one or more additional languages has become mandatory in many school systems (see *The Common European Framework of Reference for Languages*, Europa Council, 1997). Although current European recommendations encourage an early exposure to oral material in the second language (L2), experience with oral language remains relatively low in contrast to written input, resulting in limited knowledge of the pronunciation of learned words and hence, weak and inaccurate phonological representations of those words. Also, given the variation in the orthographic transparency of different writing systems, acquiring the decoding rules in an L2 can constitute a real challenge (Brysbaert, van Wijnendaele & Duyck, 2002; Commissaire, Duncan & Casalis, 2011). In this context of L2 acquisition at school, one might wonder to what extent L2 learners are able to develop and activate accurate phonological representations of the words that they read in the L2 and how these representations interact with existing L1 grapheme-to-phoneme correspondence rules.

So far, little attention has been paid to L2 phonological processes during visual word recognition by L2 learners. What is known comes mainly from research on highly-proficient adult bilinguals using techniques such as masked phonological priming which manipulate phonological overlap across languages with cognates or interlingual homographs as stimuli. These studies have shown that sub-lexical grapheme-to-phoneme correspondences (i.e., GPCs) from both languages are automatically and simultaneously co-activated during visual word recognition (Brysbaert, Van Dyck & Van de Poel, 1999; Comesaña et al., 2015; Dijkstra, Grainger & van Heuven, 1999; Duyck, Diependaele, Drieghe & Brysbaert, 2004; Gollan, Forster & Frost, 1997; Haigh & Jared, 2007; Jared & Szucs, 2002; Kim & Davis, 2003; Nakayama, Sears, Hino & Lupker, 2012; Nas, 1983; Schwartz, Kroll & Diaz, 2007;

Van Wijnendaele & Brysbaert, 2002; Zhou, Chen, Yang & Dunlap, 2010). This conclusion is consistent with the language non-selective lexical access hypothesis (Dijkstra & van Heuven, 2002) and extends the hypothesis to the sub-lexical orthography-to-phonology interface. Indeed, current models of (adult) bilingual visual word recognition such as BIA+ (Dijkstra & van Heuven, 2002), an extension of BIA (Bilingual Interactive Activation, Dijkstra, van Heuven & Grainger, 1998), and Multilink (Dijkstra et al., 2018), support the idea of automatic phonological activation during silent reading and pre-lexical co-activation of representations, across languages.

Although these mechanisms have been described in relation to highly proficient adult bilinguals (see also Sauval, Perre, Duncan, Marinus & Casalis, 2017, for findings from highly proficient/early bilingual children), not much is known about how these phonological processes develop for learners of an L2 in a school context, for whom the phonological lexicon is less developed and knowledge of L2 GPCs relatively incomplete (e.g., Commissaire et al., 2011; Valente, Ferré, Soares, Rato & Comesaña, in press; see also Poarch & van Hell, 2012, for work on speech production). In the school context, two key features need to be taken into account. Firstly, the challenges that face L2 learners during the initial steps of L2 acquisition with regard to the acquisition of L2 GPCs. For instance, in the case of French and English, some of these GPCs are shared across languages (e.g., most consonants), while many others are specific to the L2. These L2-specific GPCs may relate to graphemes that do not occur in the L1 (e.g., the grapheme ‘oa’ is specific to English; the grapheme ‘on’ corresponding to one phoneme /ɔ̃/ occurs only in French) or to correspondences that are incongruent across languages (e.g., ‘i’ is mainly pronounced as /i/ in French but as /aɪ/ or /ɪ/ in English). An exploration of this issue by Commissaire, Duncan, and Casalis (2014, Experiment 3) found that French 6th and 8th graders learning L2 English took longer to detect a letter from an English word when its phonemic correspondence was incongruent with

French GPCs (e.g., slower to detect ‘a’ in *name* than in *past*). This was interpreted as reflecting the influence of L1 GPCs on L2 word processing, consistent with language non-selectivity during phonological activation. Nevertheless, alternative interpretations such as whether the target letter was processed using its French or English letter name could not be ruled out, leading to rather limited generalisation to more ecological reading situations. To our knowledge, no other study has explored phonological processing during L2 visual word recognition or how L1 and L2 GPCs interact among child L2 learners (but see Valente et al., in press, on cognate *auditory* learning and processing by grade 5 and adult learners).

Secondly, it is important to remember that young L2 learners are also developing readers, a situation that implies the use of non-expert reading mechanisms. It seems reasonable to expect that studies of monolingual reading acquisition may be able to improve our understanding of processes by L2 school learners, and in particular, to offer insights into the role of automatic phonological activation during visual word recognition. However, despite several studies involving a variety of tasks, grade levels and writing systems, no consensus about this role has yet emerged. Masked priming studies either report automatic phonological activation throughout primary school (Booth, Perfetti & MacWhinney, 1999; Sauval, Perre & Casalis, 2017; Ziegler, Bertrand, Lété & Grainger, 2014), or else, non-significant phonological priming effects (Comesaña, Soares, Marcet & Perea, 2016; Davis, Castles & Iakovidis, 1998). This discrepancy is likely to be due to factors such as the use of different priming procedures along with differing SOAs, prime/target pairs with varying degree of orthographic overlap and the characteristics of individual writing systems (see Rastle & Brysbaert, 2006, for a discussion of phonological priming in English-speaking adults).

Other support for phonological activation during visual word recognition by developing readers comes from the pseudohomophone interference effect (Goswami, Ziegler,

Dalton & Schneider, 2001; Grainger, Lété, Bertrand, Dufau & Ziegler, 2012). Nevertheless, here also contradictory findings have emerged, possibly due to differences in grade level and writing systems. Grainger et al. (2012) observed that French children in grades 1-5 found it more difficult to reject pseudowords that are homophonic to real words (e.g., *brane*) in a lexical decision task. This pseudohomophone interference effect, which has also been shown in adults (Braun, Hutzler, Ziegler, Dambacher & Jacobs, 2009; Briesemeister et al., 2009; Ziegler, Jacobs & Klüppel, 2001) seems to decrease during reading acquisition. Goswami et al. (2001) reported a similar effect in a German-speaking sample of 8-9-year-olds but not among English-speaking children matched on chronological age. Like the previous work, this study was conducted in the participants' L1, and the language group difference was attributed to cross-linguistic differences in orthographic depth (or transparency): readers of an opaque orthography such as English rely on a direct orthographic reading procedure more than readers of transparent orthographies who rely on the one-to-one GPCs and an indirect phonological reading procedure (i.e., the Orthographic Depth Hypothesis, Frost, Katz & Bentin, 1987; see also the Psycholinguistic Grain Size theory by Ziegler & Goswami, 2006). A recent theoretical proposal by Lallier and Carreiras (2018) argues that early bilingual children who learn to read in each of their two languages adapt their reading procedures according to the degree of transparency of the language they are currently reading. However, the case of L2 word recognition, especially when the L2 is learned late and in a school context, is challenging, as reading processes may depend on the features not only of the L2 but also of the L1. Despite the extensive literature on cross-language transfer of reading and reading-related *skills* (e.g., Bialystok, Luk & Kwan, 2005; Commissaire et al., 2011), the extent to which reading *mechanisms* or *strategies* are transferred from L1 to L2 has received little attention.

The present study

The goal of the present study is to investigate whether young L2 learners automatically activate phonological information when processing the L2 in the written modality, and whether evidence can be found for co-activation of L1/L2 GPCs in this population. After reviewing the limited research on this issue in young L2 learners (, we chose to investigate the pseudohomophone interference effect as this technique seems to provide the most consistent evidence about phonological activation by developing readers (in their L1). We have adopted a developmental (cross-sectional) approach to the investigation of the learning process by examining participants with varying L2 proficiency and exposure. These participants are French-speaking adolescents attending secondary grades 6 and 8 who have been learning L2 English for only a few months and for around two years, respectively. The extent to which L1 and L2 GPCs are co-activated is investigated; in particular, whether L2 learners are sensitive to pseudohomophony with L2 words during L2 silent reading when using within-language (L2) and between-language (L1) GPCs. Hence, pseudowords have been constructed so as to be homophonic with real L2 words, either using (a) L1 GPCs (e.g., *grine* from the base word *green*) to focus on the issue of cross-language interactions at the sub-lexical level (Experiment 1); or (b) L2 GPCs (e.g., *grean*) to investigate within-language phonological mechanisms (Experiment 2, using new participants). We also examine the L2-to-L1 direction by testing whether a pseudohomophone effect emerges in an L2 task for pseudowords that are homophonic with real L1 words when using L2 GPCs (e.g., *droal* from the French base word *drôle*, /dʁol/, [*funny* in English], Experiment 3).

Experiment 1

Experiment 1 tested the pseudohomophone interference effect in an English (L2) lexical decision task by using cross-language homophony. Pseudohomophones (PsH), pseudowords that sound like real English (L2) words when using French (L1) correspondences (e.g., *grine* from the base word *green*), were contrasted with orthographic control pseudowords (OCs) that differed by one letter from the PsH (e.g., *grane*) but shared the same degree of orthographic similarity as the PsH with the base word. From the monolingual literature, we hypothesized that PsHs would be harder to reject as nonwords than OCs if participants automatically activated phonology during L2 silent reading. Yet, while significant PsH interference effects have been found in the L1 of French developing readers (Grainger et al., 2012; Sauval, Perre & Casalis, 2017; Ziegler et al., 2014), the extent to which these effects may be observed in an English task is unclear due to the previous contradictory results (Goswami et al., 2001; Frost et al., 1987; Ziegler & Goswami, 2006). Our design enabled us to test whether L1 GPCs were activated even though the task was being conducted in the L2, an outcome that would indicate the co-activation of GPCs at the sub-lexical orthography-to-phonology interface. The developmental question of whether phonological activation would differ between the two grade levels was also of interest to shed light on the dynamics of learning an L2.

Method

Participants

A total of 60 adolescents participated in this experiment including 31 participants from grade 6 (mean age = 11;9, SD = 3 months) and 29 from grade 8 (mean age = 13;8, SD = 4 months). All were native French-speakers and none of the participants had ever lived in a bilingual environment or spoken English in their daily lives. Participants were recruited from three schools in two French regions, Pays de la Loire and Grand Est. While grade 8

adolescents had been learning written English for two and a half years, grade 6 children had only around six months of English learning at secondary school. Nevertheless, most of the children had also been exposed to some English throughout primary school, although generally in its oral form. French (L1) reading skills were assessed by using the text reading task, l'Alouette (Lefavrais, 1965) which yields indices of reading accuracy (the number of correctly read words within a time limit of 180 seconds divided by the total number of words in the text) and reading speed (computed by multiplying the number of correctly read words by 180, and dividing by the actual time taken by the participant). One grade 6 child was removed from the initial sample due to poor reading skills (performance lower than 2.5 SD below the mean), leading to a final sample of 59 participants. One-tailed t-tests confirmed that grade 8 children read more accurately (accuracy index = 97.6, SD = 1.2) than grade 6 children (accuracy index = 96.9, SD = 1.4, $t(57) = 1.966$, $p < .03$) and the older children also read faster (speed index = 454, SD = 80 and 355, SD = 83, for grades 8 and 6, respectively, $t(57) = 4.719$, $p < .001$). In order to verify different levels of L2 proficiency between the two groups, a productive vocabulary test (L1-to-L2 translation, Commissaire et al., 2011) was also administered. This confirmed that the two groups differed in L2 proficiency (mean grade 6 = 39% of words correctly translated, SD = 17; mean grade 8 = 66%, SD = 17, $t(57) = 5.991$, $p < .001$). Testing took place at the end of the school year, between April and June.

Stimuli

A total of twenty high frequency English words (mean frequency = 458, SD = 347, Children's Printed Word Database, CPWD, Masterson, Stuart, Dixon & Lovejoy, 2003, see appendix A) were selected as base words for the creation of the experimental pseudoword stimuli. From these words, PsHs were created to sound the same as their base words only when using French (L1) GPCs. Note that PsHs were phonologically very similar to their

corresponding base words but not identical due to cross-language differences in phonemic repertoires. Orthographic control pseudowords (OCs) were created by changing one or several letters in the PsH. We ensured that PsHs and OCs had the same degree of orthographic similarity to their base words (e.g., *grine* and *grane* from the base word *green*) as this would imply a phonological locus for any PsH effect that emerged. The two conditions were thus matched according to the normalized Levenshtein Distance (NLD, Schepens, Dijkstra, & Grootjen, 2012 taken from NIM database, Guasch, Boada, Ferré, & Sanchez-Casas, 2013), $t < 1$, n.s. [mean PsH/base word NLD = .47, SD = .12; mean OC/base word similarity = .49, SD = .14]. Further, it was important to ensure that any difference to emerge between the two pseudoword conditions could not be due to one pseudoword being more word-like than the other in either language; in other words, it was essential to avoid the ‘markedness’ effect that is known to influence L2 written processing (e.g., Commissaire, Audusseau & Casalis, in press; Van Kesteren, Dijkstra & de Smedt, 2012). Therefore, PsHs and OCs were also matched according on English (L1) bigram frequency ($t < 1$, n.s; mean PsH = 2735, SD = 1197; mean OC = 2559, SD = 755) and French (L2) bigram frequency ($t < 1$, n.s; mean PsH = 5903, SD = 1972; mean OC = 6037, SD = 2452) using the MCWord database (Medler & Binder, 2005) and the Lexique database (New, Pallier, Ferrand & Matos, 2001), respectively). The above databases were also used to match the conditions on: French minimal bigram frequency (i.e., the frequency of the least frequent bigram in the word), $t < 1$, n.s. [mean PsH = 1617, SD = 904; mean OC = 1688, SD = 911], a relevant variable in L2 visual word recognition (Commissaire et al., in press); and English trigram frequency, word and rhyme neighborhood size and frequency (see appendix B). It is also noteworthy that none of the pseudowords were real L1 words or PsHs of real L1 words.

To minimize possible strategic effects arising from exposure to PsHs (McQuade, 1991), twenty fillers were added to the pseudoword stimuli so that only a third of the

pseudowords were PsHs (see appendix E). These fillers were created by changing one letter from an English word (*groat* derived from the base word *great*) and did not share any phonological overlap with real words. For the purpose of the lexical decision task, 60 high frequency English words (mean frequency = 588, SD = 659, CPWD, Masterson et al., 2003) were also selected.

Procedure

A fixation point was presented to the participants for 1000 ms followed by a mask composed of several hashes presented for 1000 ms. The target then appeared until the participant responded or disappeared after 3000 ms if there was no response. Participants were asked to decide as quickly and accurately as possible whether the target letter string on the screen was an English word or not.

After performing the lexical decision task, we collected participants' responses about their knowledge of the base words as real English words¹ by inserting the base words into a list which also included 20 pseudowords and asking them to circle those items that they recognized as real English words (see results below).

Results

Due to a technical problem, only the data from 47 participants (25 from grade 6 and 22 from grade 8) could be analysed. Participants' means (and standard deviations) by group and by pseudoword condition can be found in Table 1. We have included descriptive statistics on

¹ We decided against using an orthographic choice task in which participants have to choose between a correct spelling and a pseudohomophone (e.g., *bird* vs. *beurde*) as a way to investigate participants' knowledge of base word spellings. Exposing the participants to the pseudohomophones before the lexical decision task would have biased lexical decision responses, and, equally, orthographic choices would have been biased toward incorrect responses (i.e., PsHs) if completed after the lexical decision task.

all of the pseudoword data (i.e., the global analysis) and on the data from the responses corresponding to only the base words that were recognized at post-test (i.e., the partial analysis). Note that in order to understand how L2 learners at different ages and levels of L2 proficiency perform an L2 lexical decision task, word accuracy and reaction times by group are also presented. This also enables us to compare participant groups across Experiments 1 to 3 as the same target words were used across all three experiments.

All data points below 300 ms and above 3000 ms were removed from the reaction time analyses (i.e., less than 1% of word and pseudoword data). Reaction times and accuracy were respectively examined with linear mixed models and binomial mixed models using the lme4 package on R (Bates, Mächler, Bolker & Walker, 2015). For all the models considered below, we included a random intercept for participants and items². We used a model comparison approach by examining 1) the Akaike Information Criteria (AIC) according to which lower values equal better model fit to data (Akaike, 1973) and 2) p values based on likelihood ratio test comparisons.

Word data

Although the focus of the present experiment was on how L2 learners processed pseudowords, we conducted a preliminary analysis on word items to investigate grade effects. For both reaction times and accuracy data, we compared model 0 in which only random participant and item effects were considered to model 1 that also included grade as a fixed effect. As expected, the reaction time model that included grade as a fixed effect (model 1 AIC = 31368) had a better fit compared to the model that did not include it (model 0 AIC = 31376; $X^2(1, 47) = 10.39, p < .01$). A similar pattern emerged for accuracy models (AIC = 2302 and AIC = 2316 for model 1 and model 0 respectively; $X^2(1, 47) = 16.12, p < .001$).

² In word data analyses, target words were entered as a random variable. In pseudoword data analyses, base words from which PsHs and OCs were created and entered as the random variable.

Indeed, grade 8 participants were more accurate and faster than grade 6 participants when deciding that the item was a real L2 word (see Table 1), confirming the academic progress in processing L2 word items between these secondary school grades.

Pseudoword data

Pseudoword data were analysed by considering, for each participant, only the items for which the corresponding base words were known as real words (on average 70% and 90% of items for grade 6 and 8 groups, respectively) according to their individual responses at post-test. Pseudohomophones may be processed differently compared to orthographic controls only if the corresponding base word is known as a word.

Firstly, we examined the influence of pseudoword status by comparing a model (model 0) that only assumed random participant and item effects to a model (model 1) that included pseudoword status as a fixed effect. Secondly, we compared the best previous fitted model to a model that 1) also included the fixed effect of group (model 2) and 2) that added the fixed effect of group and the fixed effect of interaction between group and the other variable(s) having a significant fixed effect (model 3). Note that given the reduced number of items per participant to be considered in the reaction time analyses (especially in the grade 6 group), we also examined the pseudoword status model in each grade group to ensure that effects were not mostly due to the grade 8 group whose data were more numerous.

Reaction times

None of the tested models that included a fixed effect of pseudoword status (model 1; AIC = 18589), a fixed effect of group (model 2; AIC = 18588) or a fixed effect of interaction between status and group (model 3; AIC = 18588) provided a better adjustment compared to model 0 that only included random intercepts for participants and items (AIC = 18589).

Examining the pseudoword status model by group condition revealed a better adjustment of model 1 compared to model 0 in grade 8 adolescents ($AIC = 9824$ and $AIC = 9826$ respectively, $X^2(1) = 4.37, p = .036$), but still no difference between model 0 and 1 in grade 6 group ($AIC = 8751$ and $AIC = 8753$ respectively, $X^2 < 1$, n.s.)

Table 1. Reaction times and error rates (and standard deviations) for word and pseudoword items (including experimental pseudowords *and* fillers) in Experiment 1. Data from the experimental PsH and OC pseudowords are further broken down into a global analysis with all observations and a partial analysis based only on known base words.

	Words	Pseudowords	PsHs (global analysis)	OCs (global analysis)	PsHs (partial analysis)	OCs (partial analysis)
Grade 6 learners						
Reaction times	978 (156)	1102 (249)	1089 (245)	1117 (261)	1098 (228)	1116 (277)
Error rate	28.7 (12)	14.5 (13)	16.8 (13)	9.4 (13)	21.3 (17)	8.8 (14)
Grade 8 learners						
Reaction times	839 (157)	985 (222)	1003 (250)	969 (213)	1016 (264)	973 (223)
Error rate	14.1 (10)	12.9 (10)	14.3 (12)	10 (12)	14.8 (13)	10.4 (14)

Errors

Model 1 (AIC = 1058) had a better fit than model 0 (AIC = 1076, $X^2(1) = 20.02$, $p < .001$) suggesting an effect of pseudoword status. Indeed, participants made more errors on the PsH (16.9% errors) than the OC condition (9.7% errors). The addition of the fixed effect of group to model 1 (model 2) did not improve adjustment compared to model 1 (AIC = 1059; $X^2 < 1$, n.s.), and neither did model 3 that included a fixed effect of interaction between status and group (AIC = 1059). Note that the model comparison yielded stronger differences between model 0 and 1 in grade 6 adolescents (AIC = 554 and AIC = 537 respectively, $X^2(1) = 18.27$, $p < .001$) than in grade 8 adolescents (AIC = 537 and AIC = 535 respectively, $X^2(1) = 4.05$, $p = .044$).

Discussion

The present experiment revealed a PsH interference effect in an English (L2) lexical decision task: more errors were made for PsHs that sounded like real English words when using French (L1) GPCs than for orthographic controls. The fixed effect of interaction between pseudoword status and grade did not provide a better model adjustment; and yet, it is noteworthy that the PsH interference effect on the error data was stronger in grade 6 than in grade 8. Surprisingly, reaction time models that assumed a fixed effect of pseudoword status effect did not provide a better adjustment than models that only included random factors. Nevertheless, separate analyses by group revealed that grade 8 (but not grade 6) adolescents took longer to reject PsHs compared to OCs. This pattern of results was robust since it was not only obtained in the global analysis but also in the partial analysis, where it was clear that the base words were known.

Given that both types of pseudoword were equally orthographically similar to the base word, it is very likely that the disadvantage for PsHs reflects the homophony with the base words rather than any orthographically mediated effect. Thus, this pattern of results

demonstrates that French adolescents learning English as an L2 activate phonological information during silent reading in that L2. More specifically, Experiment 1 shows that *L1* GPCs were automatically activated as English base words (e.g., *green*) were activated through their cross-language PsHs that were created using French L1 GPCs (*grine*). Interestingly though, this PsH effect was not consistent across reaction time and error data, and the pattern of results seemed to differ slightly between the two grade groups. While grade 6 adolescents only showed a PsH interference effect on the error data, grade 8 adolescents showed this effect on both error (although to a lesser extent than the grade 6 group) and reaction time data. Before we try to interpret this pattern, it is worthwhile to mention that, unsurprisingly, grade 8 children were more proficient in the L2 than the younger grade 6 group: indeed, they performed better in response to English words in the lexical decision task (i.e., better accuracy and faster reaction times), a finding in line with their higher proficiency in the L1-to-L2 translation task. Beyond this L2 proficiency difference between the two groups, the groups also differed on their L1 reading skills as adolescents continue to improve their literacy skills throughout secondary school, and thus, the relative importance of phonological activation compared to more direct orthographic activation during silent reading may also continue to change. Further consideration of these variations will be provided in the General Discussion.

Thus, Experiment 1 showed cross- language phonological activation during L2 silent reading that was observed as early as secondary grade 6, after only a few months of L2 written exposure and learning. The strong influence of L1 GPCs in activating L2 word representations was not unexpected given their saliency when learning an L2 in a school context. Whether L2 words could also be activated by within-language PsHs was investigated in Experiment 2.

Experiment 2

As in Experiment 1, PsHs of real English (L2) words were included in the stimulus list of a lexical decision task and presented to grade 6 and grade 8 English learners. In comparison with the previous experiment, PsHs were constructed using English GPCs only (e.g., *dreem* from the base word *dream*). They were derived from base words, 65% of which had already been used in Experiment 1, and were again contrasted with orthographic controls (e.g., *droam*). We hypothesized that a PsH interference effect would reflect automatic activation of L2 specific GPCs. Whether PsH interference effect would be observed at both grade levels was also of interest in relation to the developmental dynamics of L2 GPC learning. We expected that a PsH interference effect would be shown either at both grade levels or else at grade 8 only since the older participants would be likely to have stronger L2 GPC knowledge (Commissaire et al., 2011).

Method

Participants

A total of 55 secondary school adolescents participated in this experiment, including 31 6th graders (mean age = 11;8, SD = 5 months) and 24 8th graders (mean age = 13;5, SD = 4 months). They were recruited from three schools in the French region Grand Est. As expected, grade 8 children read more accurately and faster than grade 6 children (mean accuracy index = 97.6, SD = 1.4 and 96.2, SD = 2.7 for grades 8 and 6, respectively, $t(53) = 2.41$, $p < .02$; mean speed index = 448, SD = 90 and 342, SD = 97 for grades 8 and 6, respectively, $t(53) = 4.147$, $p < .001$). Also, the L1-to-L2 translation test (See Experiment 1) revealed different knowledge of English vocabulary between 6th graders (mean = 16.6% of correctly translated words, SD = 10) and 8th graders (mean = 64.5%, SD = 20, $t(53) = 11.409$, $p < .001$). Testing took place in the middle of the school year, between January and February.

Stimuli and procedure

A total of twenty high frequency English words were selected as base words from the CPWD database (Masterson et al., 2003; see appendix C), among which 13 had already been used in Experiment 1. PsHs were created to sound like English base words using English L2 GPCs only (e.g., *dreem* from the base word *dream*). As in Experiment 1, the two pseudoword conditions were matched on their NLD with the base word, $t < 1$, n.s. [mean PsH = .58, SD = .17; mean OC = .62, SD = .15], English mean bigram frequency, $t < 1$, n.s. [mean PsH = 2371, SD = 1488; mean OC = 1987, SD = 862] and French mean bigram frequency, $t < 1$, n.s. [mean PsH = 4678, SD = 3820; mean OC = 4039, SD = 2921]. The two conditions were also matched on minimal bigram frequency according to French statistics, $t < 1$, n.s. [mean PsH = 710, SD = 1655; mean OC = 877, SD = 888] and other English sub-lexical and lexical properties (see appendix B). The same twenty fillers and sixty words used in Experiment 1 were added and the same procedure was used.

Results

Descriptive data on words and pseudowords are provided in Table 2. The procedures described for data cleaning and analysis in Experiment 1 were used again here.

Word data

Not surprisingly, a reaction time model that included grade as a fixed effect (model 1 AIC = 37750) had a better fit compared to a model that did not include it (model 0 AIC = 31752; $X^2(1) = 4.66$, $p = .03$). A similar pattern emerged for accuracy (AIC = 2772 and AIC = 2811 for model 1 and model 0 respectively; $X^2(1) = 40.69$, $p < .001$). This reflected the fact that grade 8 participants were more accurate and faster than grade 6 participants when deciding whether or not an item was a real L2 word (see Table 2).

Pseudoword data

As before, an additional partial analysis is reported based only on the data corresponding to known base words, as indicated by the individual responses collected at post-test (60% and 90% of items were known by grade 6 and 8 groups, respectively).

Reaction times

Model 1 (AIC = 17158) had a slightly better fit than model 0 (AIC = 17159, $X^2(1) = 3.61, p = .057$) suggesting a trend for an effect of pseudoword status. Participants took longer to reject PsHs (1150 ms) than OC pseudowords (1117 ms). Neither the addition of the fixed effect of group to model 1 (model 2) or the inclusion of a fixed effect of interaction effect between group and status (model 3) improved model adjustment (AIC = 17159, $X^2 < 1$, n.s. and AIC = 17160, $X^2(1) = 1.41, p = .24$, n.s. for models 2 and 3 respectively). Examining the pseudoword status model by group condition revealed a better adjustment of model 1 compared to model 0 in grade 8 adolescents (AIC = 10272 and AIC = 10274 respectively, $X^2(1) = 4.51, p = .034$), but no difference between model 0 and 1 in the grade 6 group (AIC = 6897 and AIC = 6898 respectively, $X^2 < 1$, n.s.).

Errors

Model 1 (AIC = 1608) had a better fit than model 0 (AIC = 1647, $X^2(1) = 40.53, p < .001$) suggesting an effect of pseudoword status. Indeed, participants made more errors on PsHs (31.4% errors) than OC pseudowords (18.8% errors). The addition of the fixed effect of group to model 1 (model 2) improved adjustment compared to model 1 (AIC = 1585; $X^2(1) = 25.12, p < .001$), showing that grade 6 adolescents made more errors (35.2 %) than grade 8

(16.3 %). Model 3 ($AIC = 1587$, $X^2 < 1$, n.s.) did not improve the adjustment; PsH effects were 13.6% and 11% for grade 6 and 8 adolescents, respectively.

Table 2. Reaction times and error rates (and standard deviations) for word and pseudoword items (including experimental pseudowords *and* fillers) in Experiment 2. Data from the experimental PsH and OC pseudowords are further broken down into a global analysis with all observations and a partial analysis based only on known base words.

	Words	Pseudowords	PsHs	OCs	PsHs	OCs
			(global analysis)		(partial analysis)	
Grade 6 learners						
Reaction times	1040	1138	1164	1125	1153	1139
	(260)	(230)	(246)	(242)	(304)	(270)
Error rate (%)	27.8	31.3	37.4	26.6	42.6	29
	(10)	(15)	(16)	(17)	(19)	(19)
Grade 8 learners						
Reaction times	910	1093	1130	1076	1146	1089
	(211)	(239)	(261)	(248)	(268)	(257)
Error rate (%)	10.7	15.6	21.5	11.7	22.1	11
	(6)	(9)	(11)	(10)	(12)	(11)

Discussion

Experiment 2 yielded a significant PsH interference effect on both reaction time and error data which indicated that L2 learners automatically activate L2-specific GPCs during L2

silent word reading. Thus, in addition to the activation of L1 GPCs observed in Experiment 1, L2 learners can also activate within-L2 sub-lexical units when reading in the L2.

Interestingly, although the observed PsH interference effect was numerically comparable across the two groups in the error data, the PsH effect on the reaction time data was mostly driven by the grade 8 group, as in Experiment 1³.

Before attempting to understand the mechanisms underlying these effects in L2 learners, some alternative hypotheses need to be considered. In both experiments, PsH were compared to OCs that shared the same degree of orthographic similarity with base words to ensure that any effect would reflect the phonological influence of the PsHs. However, there was considerable heterogeneity in the degree of orthographic similarity between Experiments 1 and 2 as well as within each experiment. Indeed, orthographic similarity was much higher in Experiment 2 compared to Experiment 1 due to the manipulation of within- language GPCs in the former (e.g., *beurde* vs. *burd* for the base word *bird* in Experiment 1 and 2 respectively); this, together with the slight differences in the proficiency levels of the grade 6 participants, seems likely to explain the higher number of false alarms in Experiment 2⁴ ().

The degree of orthographic similarity between PsHs and their base words also varied within each experiment and the impact of this variable needs to be examined to shed further light on underlying mechanisms. In fact, it is likely that the more the orthographic overlap

³ Importantly, performances in the post- test measuring knowledge of base words were very close across Experiments 1 and 2, for each grade level, enabling comparison between these two experiments. Nevertheless, the grade 6 participants in Experiment 1 had a larger L2 productive vocabulary than those in Experiment 2 [$t(54) = 6.55, p < .001$], as measured by the L1-to-L2 translation task, most likely due to a later testing time during the academic year. Grade 8 participants had similar L2 proficiency across the two experiments [$t < 1$, n.s.].

⁴ Note that all experimental pseudowords from Experiment 2 also had lower French minimal bigram frequency than items from Experiment 1, which is likely to lead to a stronger orthographic markedness effect, that is, a general higher rate of pseudoword errors (see Commissaire et al., in press).

between a PsH and its corresponding base word, the stronger the PsH interference effect due to the greater competition between orthographic and phonological lexical representations. We therefore conducted a follow-up analysis on the error data from Experiments 1 and 2 (using known base words only) by entering orthographic similarity (OS) for each PsH as a continuous variable. Adding this control variable in all the compared models yielded a similar pattern of results as previously described. Detailed methodological and statistical information can be found in appendix F. (We also examined the extent to which OS would interact with the PsH interference effect by comparing the best fitted model to a model that also included a fixed effect of interaction between OS and pseudoword status. No greater adjustment was found for this new model in Experiment 2. Despite greater adjustment for this new model in Experiment 1, examination of PsH effects according to OS measures revealed that PsH effects were actually slightly stronger for the two items with lowest OS indices, a finding that is counter to what would be predicted and so is likely to reflect an item- related artifact (e.g. related to base word frequency).

One alternative interpretation could be that PsHs, especially those with a higher OS, could have been mistaken for real base words (leading to a false alarm), as a result of poor L2 orthographic representations. Indeed, if L2 learners had erroneous L2 word form representations, it would be likely that this would give rise to more errors for PsHs with high OS (e.g., *grean*) than for those with lower OS (e.g., *fude*). Nevertheless, although this interpretation cannot be excluded, it would seem unlikely as all analyses were conducted on participants' known base words, that is, words that had been correctly identified as real English words among a set of distractors.

Thus, Experiments 1 and 2 showed that English (L2) words could be activated through both cross-language (L1) and within-language (L2) GPCs. Experiment 3 aimed to further examine cross-language phonological interactions by testing the influence of L2 GPCs on the

activation of L1 words. The design also enabled further testing of the phonological basis to the PsH interference effect in order to exclude the alternative hypothesis of poor L2 orthographic representations.

Experiment 3

In Experiment 3, French children performed an L2 lexical decision task containing PsHs of real French (L1) words rather than the PsHs of English words that were used in previous experiments. Importantly, these PsHs were built on the basis of L2 GPCs only (e.g., *droal*, homophone of the French word *drôle* /dʁol/, [funny in English] according to English GPC rules). As in Experiment 1, we hypothesized that a PsH interference effect would provide evidence for the existence of strong interactions between L1 and L2 lexical and sub-lexical levels of processing during word recognition. It would also confirm the phonological locus of the effect as it was highly unlikely that the PsHs that we used would correspond to erroneous L1 orthographic representations.

Method

Participants

The sample comprised 61 participants, including 39 grade 6 adolescents (mean age = 12;1, SD = 3) and 22 grade 8 adolescents (mean age = 13;11, SD = 3). They were recruited from two schools in the French Centre and Grand Est regions. Grade 8 participants read French more accurately and faster (accuracy index = 98.1, SD = 1.4; speed index = 432, SD = 65) than grade 6 adolescents (accuracy index = 96.2, SD = 3.7; speed index = 341, SD = 71). They were also more proficient in English as confirmed by the L1-to-L2 translation task

(mean correct = 62%, SD = 13 and 25%, SD = 8 for grade 8 and grade 6 participants, respectively). Testing took place between March and July.

Stimuli and procedure

A total of twenty high frequency French words were selected from the Lexique database (New et al., 2001; see appendix D). PsHs that were created sounded like their corresponding base word when using English GPCs only (e.g., *droal*, from the base word *drôle* /d^{ro}l/, [*funny* in English]). PsH and OC conditions were again matched on their NLD with the base word, $t < 1$, n.s. [mean PsH = .51, SD = .23; mean OC = .51, SD = .23], French mean bigram frequency, $t(38) = 1.113$, $p = .27$, n.s [mean PsH = 5064, SD = 3298; mean OC = 3985, SD = 2816] and minimal bigram frequency, $t < 1$, n.s. [mean PsH = 853, SD = 1330; mean OC = 1114, SD = 1020]. Filler words and procedure were the same as in the two previous experiments.

Results

Descriptive data are provided in Table 3. The same data cleaning and analyses procedures were used as in both previous experiments.

Word data

Not surprisingly, a reaction time model that included grade as a fixed effect (model 1 AIC = 45227) had a better fit compared to a model that did not include it (model 0 AIC = 45233; $X^2(1) = 7.82$, $p < .01$). The accuracy model that included a fixed effect of grade also had a better fit compared to the model that only include random variables (AIC = 2371 and

AIC = 2395 for model 1 and model 0 respectively; $X^2(1) = 26.16, p < .001$). This confirmed the academic progress in L2 learning between secondary school grades 6 and 8 (see Table 3).

Pseudoword data

The analysis was conducted on the raw data (i.e., previously referred to as the global analysis) as no post-test of base word recognition was conducted given that PsHs were constructed from real French (L1) words that should all be known to the adolescents.

Reaction times

Model 1 (AIC = 29272) had a better fit than model 0 (AIC = 23274, $X^2(1) = 3.87, p < .05$) suggesting an effect of pseudoword status. Participants took longer to reject PsHs (1124 ms) than OC pseudowords (1091 ms). Neither the addition of the fixed effect of group to model 1 (model 2) or the inclusion of a fixed effect of interaction between group and status (model 3) improved model adjustment (AIC = 29272, $X^2(1) = 1.82, p = .18$, n.s., and AIC = 2927, $X^2(1) = 1.71, p = .19$, n.s., for models 2 and 3, respectively). Note that the model including a fixed effect of pseudoword status had a better fit than a model 0 that did not include it in the grade 8 group (model 0 AIC = 19043; model 1 AIC = 19038, $X^2(1) = 6.24, p < .02$) but the two models did not differ from each other in the grade 6 group (model 0 AIC = 10238; model 1 AIC = 10240, $X^2 < 1$, n.s.).

Errors

Model 1 (AIC = 1179) had a better fit than model 0 (AIC = 1807, $X^2(1) = 29.44, p < .001$) suggesting an effect of pseudoword status. Indeed, participants made more errors on PsHs (19.6% errors) than OC pseudowords (12.4% errors). Adding the fixed effect of group to model 1 did not significantly improve the model 2 adjustment (AIC = 1178; $X^2(1) = 3.05, p$

=.08), despite a trend for lower accuracy for grade 6 (19.6% errors) compared to grade 8 (14% errors) participants. Model 3 (AIC = 1780) did not improve the adjustment: indeed, similar PsH effects were found in the two groups (7.7% and 6.8% difference for grades 6 and 8, respectively).

Table 3. Reaction times and error rates (and standard deviations) for word and pseudoword items (including experimental pseudowords *and* fillers) in Experiment 3. Data from the experimental PsH and OC pseudowords are subject to a global analysis only (including all observations).

	Words	Pseudowords	PsHs (global analysis)	OCs
Grade 6 learners				
Reaction times	1022 (147)	1189 (235)	1169 (250)	1162 (232)
Error rate (%)	22.6 (8.8)	20.7 (13.9)	23.4 (16)	15.7 (13)
Grade 8 learners				
Reaction times	923 (152)	1129 (202)	1113 (223)	1069 (200)
Error rate (%)	9.5 (7.6)	15.6 (15.4)	17.4 (19)	10.5 (13)

Discussion

In Experiment 3, French L2 learners showed a PsH interference effect, making more errors with pseudowords that sounded like French (L1) words, despite these being constructed with English (L2) GPCs. This pattern was the same even when the degree of orthographic similarity between PsHs and their base words was controlled for in the analyses (see Appendix F). This outcome is consistent with the results of Experiment 1 and provides further evidence for the existence of strong cross-language interactions at the sub-lexical phonological level, a finding only previously observed with highly-proficient adult bilinguals (e.g., Brysbaert et al., 1999). Thus, despite L2 GPCs being acquired relatively recently by the two groups of L2 learners, these correspondences were activated and enabled a connection to be made to L1 phonological representations.

These findings also reinforce a phonological interpretation of the L2 PsH interference effects found in Experiments 1 and 2, as opposed to the alternative hypothesis that these effects had an orthographic locus due to incorrect L2 orthographic representations. There would be no reason for these French adolescents to have incorrect L1 lexical orthographic representations and so a phonological explanation of the PsH interference effect is most likely.

General Discussion

The present study is the first investigation to use the pseudohomophone (PsH) interference effect to examine both GPC use and phonological activation during visual word recognition by young learners of an L2. The PsH interference effect was significant across all experiments, supporting automatic activation of phonology during L2 word reading. Further, the PsH interference effect was significant regardless of whether pseudohomophony was created using cross-language (L1) GPCs (Experiments 1 and 3) or within-language (L2) GPCs

(Experiment 2). It is important to highlight that L2 proficiency was comparable for each grade across the three experiments, as shown by word accuracy scores on the L2 lexical decision task, the post- test performance (in Experiments 1 and 2) and their L1-to-L2 vocabulary scores (despite slight differences between Experiments 1 and 2 at grade 6 most likely due to time of testing). Thus, for the first time, we have been able to demonstrate that young L2 learners automatically activate L2 word phonology during silent reading despite their limited L2 vocabulary and GPC knowledge. Importantly, this phonological activation was found cross-linguistically, consistent with the language non-selective hypothesis. These data therefore demonstrate that adolescents in the early phases of L2 acquisition are already making use of visual word recognition mechanisms that are similar to those observed among highly proficient adult bilinguals.

Phonological activation in L2 learners

Despite the predominance of written language and poor oral language skills in the context of L2 school learning, L2 learners automatically activated phonological information during the silent reading of L2 words⁵This is in line with current models of skilled visual word recognition that assume automatic activation of phonological codes when reading in the dominant language (e.g., Perry, Ziegler & Zorzi, 2010) or in the nondominant one (e.g., Dijkstra & van Heuven, 2002), and with more recent data on young developing readers (Goswami et al., 2001; Grainger et al., 2012). The PsH interference effect is commonly interpreted as reflecting the competition between phonological and orthographic information: while PsHs activate lexical phonological forms of real base words, no corresponding lexical

⁵ For converging evidence on the role of phonology among readers with underspecified phonological representations, see Mayberry, del Giudice and Lieberman's (2010) work on participants with hearing difficulties.

orthographic candidates are activated, and this conflict leads to greater false alarms and/or slower rejection times (e.g., Braun et al., 2009; Briesemeister et al., 2009).

Interestingly, our data always yielded significant PsH interference effects in the error data but not systematically in the reaction time data: participants made more false alarms to PsHs compared to OCs and, for more advanced L2 learners, also took longer to reject them. However, although a fixed effect of interaction between group and pseudoword status never improved model adjustment compared to a fixed effect of pseudoword status, it is noteworthy that the PsH effect on reaction time data was significant in the grade 8 group in all three experiments, but never in the grade 6 group (while both participant groups made more errors with PsHs than OCs). It is currently difficult to provide an appropriate interpretation since the two groups differed on two confounded variables: L2 proficiency and L1 reading level. The pattern described above could reflect the use of slightly different mechanisms by the two groups due to their varying L2 proficiency (i.e., L2 vocabulary and/or GPC knowledge), or due to different general reading processes (i.e., varying use of phonological activation compared to more direct orthographic access; see, for example, Duñabeitia, Ivaz & Casaponsa, 2015, and Valente et al., in press, with regard to age effects on cognate processing and learning). Caution should also be exercised as the different effect sizes across the two age groups could also be an artefact of their different general cognitive abilities. A further issue is a reduced number of observations (e.g., a low number of reaction time observations after error and unknown base word data removal, especially in the grade 6 group), and hence, lower statistical power. Future studies with larger sample of participants at these two grade levels will enable further investigation of these developmental trends.

Nevertheless, our data indicate that even L2 learners with rather low L2 exposure and proficiency (i.e., from grade 6) quickly activate whole-word phonology instead of using a more sequential and rather slow decoding procedure, as beginning readers would do (Ehri,

2002; Frith, 1985; Grainger & Ziegler, 2011). The PsH interference effect was found in English, a language considered to have an opaque orthography. It is noteworthy that previous work by Goswami and colleagues (2001) on young English readers failed to find PsH interference effects in their L1, a finding that was thought to reflect a weaker reliance on the phonological code in this language (Frost, Katz & Bentin, 1987) and the use of larger units than GPCs during phonological activation (Lallier & Carreiras, 2018; Ziegler & Goswami, 2006). This would therefore suggest that learning to read in a novel language takes advantage of the mechanisms already developed in the first language, French in the case of the present study, a hypothesis referred to as cross-language transfer (Bialystok et al., 2005). Our findings are in line with previous work by Commissaire (2012), who found PsH interference effects in English children learning French as an L2, both within- L2 and across-languages. It was beyond the scope of the present study to determine whether different grain sizes are used during phonological decoding in these two languages, as suggested by the psycholinguistic grain size theory (Ziegler & Goswami, 2006), and the extent to which reading mechanisms from one language transfer to the other.

Interestingly, PsH interference effects were found for L2 pseudohomophones that were constructed based on L1 GPCs (e.g., *grine* from *green*; Experiment 1) or L2 GPCs (e.g., *grean*, Experiment 2). It should be noted that PsHs were always homophonic to real words when using GPCs from one language only; that is, *grine* is homophonic to *green* when using French GPCs only, and not English ones, and *grean* sounds like *green* according to English GPCs only. Thus, the L2 learners in this study were not only sensitive to the influence of GPCs within their L2 but also to the influence of GPCs from their L1, despite their L1 not being the target language of the task. Importantly, Experiment 3 revealed that L1 words could also be activated by pseudohomophones that were constructed by using L2 GPCs (e.g., *droal* from the base word *drôle*, [*funny* in English]). Nevertheless, the matching between PsHs and

OCs was slightly less robust in this experiment as PsHs had a greater English neighbourhood size (at the rhyme and word level) compared to OCs. However, given the rudimentary vocabulary of our participants, this is unlikely to explain the pattern of results, especially for the accuracy data.

Together these data suggest that L2 learners automatically activate print-to-sound mappings from both of their languages, a result in line with the findings of cross-language phonological priming effects in highly-proficient adult bilinguals (e.g., Brysbaert et al., 1999; Duyck et al., 2001; Gollan et al., 1997; Nakayama et al., 2012; Van Wijnendaele & Brysbaert, 2002). It also more generally supports the language non-selective hypothesis that assumes that lexical and sub-lexical information from both languages is co-activated in the initial steps of lexical access (Dijkstra & van Heuven, 2002; Dijkstra et al., 2018) but extends this work to the development of young learners of an L2. Future studies that use a longitudinal approach will help to understand how the strength of L1 versus L2 sub-lexical print-to-sound information may vary with exposure and proficiency and, thus, add further important constraints on modelling L2 reading acquisition.

Developmental models of L2 visual word recognition

The present study revealed automatic phonological activation during L2 processing by young learners of an L2 in a school context, as well as co-activation of L1 and L2 GPCs. In spite of the constraints on the present study (e.g., only twenty PsH items⁶), our findings add

⁶ Only 20 PsHs were used in each experiment of the present study, along with their 20 corresponding orthographic controls. This rather small number of experimental items per condition is due to the limited vocabulary of L2 learners (see post-test scores) and the severe constraints during stimulus selection (i.e., constructing PsHs that are homophonic to the base words when using the GPCs of one language only). Note that

important information to the limited existing knowledge base about visual word recognition during L2 learning. As far as we can establish, only two previous studies have endeavored to examine L2 visual word recognition in this population, and have focused on orthographic mechanisms, either lexical (Brenders, van Hell & Dijkstra, 2011) or sub-lexical (Commissaire, Duncan & Casalis, 2014). For instance, Brenders et al. (2011) showed that cognate words were processed faster than monolingual control words in Dutch grade 5 to 9 children learning English as an L2, a finding also commonly reported in adult highly proficient bilinguals and taken as evidence for language non-selective lexical access (e.g., Dijkstra et al., 1999). Sub-lexical orthographic coding in English as an L2 was assessed by Commissaire et al. (2014) with French children in grades 6 and 8). The authors revealed grapheme complexity effects in an L2 letter detection task and, thus, showed that graphemes were processed as perceptual units in L2. Although these effects were observed from the initial stages of L2 acquisition (from grade 6), the data revealed that graphemes involved different processing costs according to whether they were shared across languages (e.g., ‘ou’ occurs as a unit in both English and French) or specific to the L2 (e.g., ‘oa’). Phonological processes were also investigated by the same authors (Experiment 3), who found that letter detection times were faster when GPCs were congruent across languages (e.g., ‘a’ in *black*) as compared to when they were incongruent (e.g., ‘a’ in *plane*); once again consistent with language non-selectivity during activation of GPCs.

As a whole, these data suggest comparable word recognition mechanisms in L1 and L2, either orthographic or phonological, even in young learners of an L2 who cannot be considered to be expert readers. The data also support the hypothesis of language non-selectivity as a core mechanism of visual word recognition in L2 that appears also apply to L2

we also chose to include another set of 20 filler pseudowords to decrease the proportion of PsHs (preventing any phonologically-related strategies).

speakers with limited exposure and rather low proficiency in the L2 (see van Hell & Tanner, 2012, for a review of proficiency effects). This small but growing knowledge base constrains how learning dynamics may be implemented in models of L2 visual word recognition. Two important models have been proposed to account for the developmental dynamics during L2 learning: the Revised Hierarchical Model (RHM, Kroll & Stewart, 1994) and BIA-d (Grainger, Midgley & Holcomb, 2010); yet, neither model has examined phonological issues in any detail as the focus of these contributions was instead on the relationship between L1/L2 word forms (orthographic forms in particular for BIA-d) and their corresponding concepts. Moreover, the authors explicitly stated that these proposals should apply to adult L2 learning rather than to the development of child learners. Thus, our study extends current knowledge of L2 word processing with regard to phonological coding, in particular, among young L2 learners and should add important constraints on future theoretical proposals about L2 visual word recognition.

References

- Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle. In B. N. Petrov & F. Caski (Eds.), *Proceedings of the Second International Symposium on Information Theory* (pp. 267–281). Budapest: Akademiai Kiado.
- Bates, D., Mächler, M., Bolker, B. M., & Walker, S. C. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67(1), 1-48.
- Bialystok, E., Luk, G., & Kwan, E. (2005). Bilingualism, Biliteracy, and Learning to Read: Interactions among Languages and Writing Systems. *Scientific Studies of Reading*, 9, 43-61.
- Booth, J. R., Perfetti, C. A., & MacWhinney, B. (1999). Quick, automatic, and general activation of orthographic and phonological representations in young readers. *Developmental Psychology*, 35(1), 3- 19.

- Braun, M., Hutzler, F., Ziegler, J. C., Dambacher, M., & Jacobs, A. M. (2009). Pseudohomophone effects provide evidence of early lexico-phonological processing in visual word recognition. *Human Brain Mapping, 30* (7), 1977- 1989.
- Brenders, P., van Hell, J. G., & Dijkstra, T. (2011). Word recognition in child second language learners: Evidence from cognates and false friends. *Journal of Experimental Child Psychology, 109* (4), 383-396.
- Briesemeister, B. B., Hofmann, M. J., Tamm, S., Kuchinke, L., Braun, M. M., & Jacobs A. M. (2009). The pseudohomophone effect: Evidence for an orthography- phonology conflict. *Neuroscience Letters, 455*, 124-128.
- Brysbaert, M., & Duyck, W. (2010). Is it time to leave behind the Revised Hierarchical Model of bilingual language processing after 15 years of service? *Bilingualism: Language and Cognition, 13*(3), 359- 371.
- 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*, 137- 148.
- Brysbaert, M., Van Wijnendaele, I., & Duyck, W. (2002). On the temporal delay assumption and the impact of non-linguistic context effects. *Bilingualism: Language and Cognition, 5*, 199- 201.
- Comesaña, M., Ferré, P., Romero, J., Guash, M., Soares, A. P., & García-Chico, T. (2015). Facilitative effect of cognate words vanishes when reducing the orthographic overlap: the role of stimuli list composition. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 41*(3), 614- 635.
- Comesaña, M., Soares, A. P., Marcet, A., & Perea, M. (2016). On the nature of consonant/vowel differences in letter position coding: Evidence from developing and adult readers. *British Journal of Psychology, 107*(4), 651- 674.

- Commissaire, E. (2012). *Orthographic and phonological coding during L2 visual word recognition in L2 learners: lexical and sublexical mechanisms* (Doctoral dissertation). Retrieved from <https://tel.archives-ouvertes.fr/tel-00865011>.
- Commissaire, E., Audusseau, J., & Casalis, S. (in press). Disentangling cross-language orthographic neighborhood from markedness effects during L2 visual word recognition. *Psychonomic Bulletin & Review*.
- Commissaire, E., Duncan, L. G., & Casalis, S. (2011). Cross-language transfer of orthographic processing skills. A study of French children who are learning English at school. *Journal of Research in Reading*, 34(1), 59-76.
- Commissaire, E., Duncan, L. G., & Casalis, S. (2014). Grapheme coding in L2: how do L2 learners process new graphemes? *Journal of Cognitive Psychology*, 26(7), 725- 739.
- Davis, C., Castles, A., & Iakovidis, E. (1998). Masked homophone and pseudohomophone priming in children and adults. *Language and Cognitive Processes*, 13, 625-651.
- Dijkstra, T., Grainger, J. & Van Heuven, W. J. B. (1999). Recognition of cognates and interlingual homographs: The neglected role of phonology. *Journal of Memory and Language*, 41, 496-518.
- Dijkstra, A., & Van Heuven, W. J. B. (2002). The architecture of the bilingual word recognition system: From identification to decision. *Bilingualism: Language and Cognition*, 5, 175-197.
- Dijkstra, T., Van Heuven, W. J. B., & Grainger, J. (1998). Simulating cross-language competition with the bilingual interactive activation model. *Psychologica Belgica*, 38, 177-196.
- Dijkstra, T., Wahl, A., Buytenhuijs, F., Van Halem, N., Al-Jibouri, Z., De Korte, M., & Rekké, S. (2018). Multilink: a computational model for bilingual word recognition and word translation. *Bilingualism: Language and Cognition*.

- Duñabeitia, J. A., Ivaz, L., & Casaponsa, A. (2016). Developmental changes associated with cross-language similarity in bilingual children. *Journal of Cognitive Psychology*, 28(1), 16-31.
- Duyck, W., Drieghe, D., Diependaele, K., & Brysbaert, M. (2004). The size of the cross-lingual masked phonological priming effect does not depend on second language proficiency. *Experimental Psychology*, 51, 1-9.
- Ehri, L. C. (2002). Phases of acquisition in learning to read words and implications for teaching. *British Journal of Educational Psychology: Monograph Series*, 1, 7–28
- Frost, R., Katz, L., & Bentin, S. (1987). Strategies for visual word recognition and orthographical depth: A multilingual comparison. *Journal of Experimental Psychology: Human Perception and Performance*, 13, 104-115.
- Gollan, T. H., Forster, K. I., & Frost, R. (1997). Translation priming with different scripts: Masked priming with cognates and non-cognates in Hebrew English bilinguals. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, 23, 1122-1139.
- Goswami, U., Ziegler, J. C., Dalton, L., & Schneider, W. (2001). Pseudohomophone Effects and Phonological Recoding Procedures in Reading Development in English and German, *Journal of Memory & Language*, 45, 648-664.
- Grainger, J., Lété, B., Bertrand, D., Dufau, S., & Ziegler, J. C. (2012). Evidence for multiple routes in learning to read. *Cognition*, 123, 280-292.
- Grainger, J., Midgley, K. J., & Holcomb, P. J. (2010). Re-thinking the bilingual interactive-activation model from a developmental perspective (BIA-d). In M. Kail & M. Hickmann (eds.), *Language acquisition across linguistic and cognitive systems*, pp. 267- 284. Philadelphia, PA: John Benjamins.
- Grainger, J., & Ziegler, J. C. (2011). A dual-route approach to orthographic processing. *Frontiers in Psychology*, 2, 54.

- Haigh, C. A., & Jared, D. (2007). The activation of phonological representations by bilinguals while reading silently: Evidence from interlingual homophones. *Journal of Experimental Psychology: Learning, Memory & Cognition*, 33, 623- 644.
- Jared, D., & Szucs, C. (2002). Phonological activation in bilinguals: Evidence from interlingual homograph naming. *Bilingualism: Language and Cognition*, 5, 225- 239.
- Kim, J., & Davis, C. (2003). Task effects in masked cross-script translation and phonological priming. *Journal of Memory and Language*, 49, 484-499.
- Kroll, J. F., & Stewart, E. (1994). Category interference in translation and picture naming: Evidence for asymmetric connections between bilingual memory representations. *Journal of Memory and Language*, 33, 149–174.
- 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.
- Language Learning in Scotland A 1+2 Approach: report and recommendations by the Languages Working Group promotted by the Scottish government. (2011). <http://www.gov.scot/resource/0039/00393435.pdf>. Accessed 23rd January 2018.
- Lefavrais, P. (1965). *Test de l'Alouette (revision 2005)*. Paris : ECPA.
- Masterson, J., Stuart, M., Dixon, M., & Lovejoy, S. (2003). The Children's Printed Word Database. www.essex.ac.uk/psychology/cpwd.
- McQuade, D. (1981). Variable reliance on phonological information in visual word recognition. *Language & Speech*, 24, 99-109.
- Medler, D. A., & Binder, J. R. (2005). MCWord: An On-Line Orthographic Database of the English Language. Retrieved on <http://www.neuro.mcw.edu/mcword/>

- 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.
- Nas, G. (1983). Visual word recognition in bilinguals: Evidence for a cooperation between visual and sound based codes during access to common lexical store. *Journal of Verbal Learning and Verbal Behavior*, 22, 526-534.
- New, B., Pallier, C., Ferrand, L., & Matos, R. (2001). Une base de données lexicales du français contemporain sur Internet : Lexique. *L'Année psychologique*, 101, 447- 462.
- Perry, C., Ziegler, J. C., & Zorzi, M. (2010). Beyond single syllables: Large-scale modeling of reading aloud with the Connectionist Dual Process (CDP++) model. *Cognitive Psychology*, 61(2), 106-151.
- Poarch, G. J., & van Hell, J. G. (2012). Cross-language activation in children's speech production: Evidence from second language learners, bilinguals, and trilinguals. *Journal of Experimental Child Psychology*, 111, 419- 438.
- Rastle, K., & Brysbaert, M. (2006). Masked phonological priming effects in English: are they real? Do they matter? *Cognitive Psychology*, 53(2), 97- 145.
- Sauval, K., Perre, L., & Casalis, S. (2017). Automatic activation of phonological code during visual word recognition in children: a masked study in grades 3 and 5. *Reading and Writing: an interdisciplinary Journal*, 30(1), 51-67.
- Sauval, K., Perre, L., Duncan, L. G., Marinus, E., & Casalis, S (2017). Automatic phonological activation during visual word recognition in bilingual children: a cross-language masked priming study in grades 3 & 5. *Journal of Experimental Child Psychology*, 154, 64-77.

- Schwartz, A. I., Kroll, J. F., & Diaz, M. (2007). Reading words in Spanish and English: Mapping orthography to phonology in two languages. *Language and Cognitive Processes*, 22, 106-129.
- Seymour, P. H. K., Aro, M., & Erskine, J. M. (2003). Foundation literacy acquisition in European orthographies. *British Journal of Psychology*, 94, 143–174.
- The Common European Framework of Reference for Languages, Europa Council. (1997). <https://rm.coe.int/16802fc1bf>. Accessed 23rd January 2018.
- Valente, D. R., Ferré, P., Soares, A. P., Rato, A., & Comesaña, M. (in press). Does phonological overlap of cognate words modulate cognate acquisition and processing in developing and skilled readers? *Language Acquisition*
- Van Hell, J. G., & Tanner, D. (2012). Second language proficiency and cross-language lexical activation. *Language Learning*, 62, 148- 171.
- Van Kesteren, R., Dijkstra, T., & de Smedt, K. (2012). Markedness effects in Norwegian–English bilinguals: Task-dependent use of language-specific letters and bigrams. *The Quarterly Journal of Experimental Psychology*, 65, 2129- 2154.
- Van Orden, G. C. (1987). A rows is a rose: spelling, sound and reading. *Memory and Cognition*, 15, 181-198.
- Van Wijnendaele, I., & Brysbaert, M. (2002). Visual word recognition in bilinguals: Phonological priming from the second to the first language. *Journal of Experimental Psychology: Human Perception and Performance*, 28, 616- 627.
- Westbury, C., & Buchanan, L. (2002). The probability of the least likely non-length controlled bigram affects lexical decision reaction times. *Brain and Language*, 81(1-3), 66-78.

- Zhou, H., Chen, B., Yang, M., & Dunlap, S. (2010). Language nonselective access to phonological representations: Evidence from Chinese-English bilinguals. *The Quarterly Journal of Experimental Psychology*, 63, 2051–2066.
- Ziegler, J. C., Bertrand, D., Lété, B., & Grainger, J. (2014). Orthographic and phonological contributions to reading development: tracking developmental trajectories using masked priming. *Developmental Psychology*, 50(4), 1026- 1036.
- Ziegler, J. C., & Goswami, U. (2006). Reading acquisition, developmental dyslexia, and skilled reading across languages: a psycholinguistic grain size theory. *Psychonomic Bulletin*, 131(1), 3- 29.
- Ziegler, J. C., Jacobs, A. M., & Kluppel, D. (2001). Pseudohomophone effects in lexical decision: Still a challenge for current word recognition models. *Journal of Experimental Psychology: Human Perception and Performance*, 27 (3), 547- 559.

Appendix A.

English base word, *pseudohomophone/orthographic control* used in Experiment 1

Bird, *beurde/baurde*; Blue, *blou/blie*; Care, *quaire/quaure*; Clean, *cline/clune* ; Cook, *couc/couf* ; Dream, *drime/drome* ; Fear, *firre/furre*; First, *feurst/faurst*; Food, *foude/foide*; Good, *goude/goide*; Green, *grine/grane*; Keep, *quipe/quape*; Kill, *quile/quile*; Moon,

moune/monn, Need, *nide/nade*; Run, *reune/roune*, Shoes, *chouse/chonse*; Should, *choude/thoud*; Slow, *slaut/slart*; Snow, *snaud/snaid*.

Appendix B.

Features of the two experimental pseudoword conditions: measures of English trigram frequency, neighborhood size and mean frequency at both rhyme and word levels.

	PsH	OC
Experiment 1		
Trigram frequency	241 (224)	222 (132)
Rhyme N size	2.8 (3.8)	2.7 (3.6)
Rhyme N mean frequency	333 (1193)	184 (610)
Word N size	2.5 (3.2)	3 (3.2)
Word N mean frequency	24 (42)	24 (34)

Experiment 2

Trigram frequency	173 (148)	152 (127)
Rhyme N size	4.3 (3.6)	3.1(3.2)
Rhyme N mean frequency	193 (413)	252 (792)
Word N size	5.4 (5.1)	4.8 (4.1)
Word N mean frequency	127 (282)	63 (138)

Experiment 3

Trigram freq	180 (127)	125 (141)	
Rhyme N size	9.2 (7.5)	6.1 (9)	
Rhyme N mean frequency	263 (401)	85 (138)	‘
Word N size	8.6 (6.3)	5.1 (4.3)	*
Word N mean frequency	49 (47)	33 (56)	

N.B. * $p = .047$; ‘ Rhyme mean frequency : $p = .068$; ‘

Appendix C.**English base word, pseudohomophone/orthographic control used in Experiment 2**

Bird, *burd/bard*; Blue, *bloo/bloi*; Dream, *dreem/droam* ; Fear, *fere/feam*; First, *furst/farst*;
Food, *fude/fope*; Game, *ghaim/gaum*; ghost, *goast/goist*; Girl, *gearl/goarl*; Good,
gould/goard; Green, *grean/greun*; Hope, *hoap/hoip*; Joke, *joak/jook*; Moon, *mewn/mawn*,

Night, *knite/snite*; Rain, *rane/raip*; Shoes, *shoose/shoise*; Should, *shood/shoad*; Slow, *sloe/sloy*; Snow, *snoe/snou*.

Appendix D.

French base word (phonetic translation; English translation), *pseudohomophone/orthographic control* used in Experiment 3

Lit (/li/; bed) *lea/loa*; Vite (/vit/; quick) *veat/voat*; vide (/vid/; empty) *veed/voed*; Lire (/lir/; read) *lear/loar*; Ville (/vil/; town) *veel/voel*; Frère (/frɛr/; brother) *frare/frore*; Oreille (/ɔʁɛj/; ear) *oreigh/oreith*, Semaine (/səmɛn/; week) *semane/semine*; Vert (/vɛr/; green) *vare/vore*; Midi (/midi/; midday) *meady/moady*; Rouge (/ruz/; red) *rooge/roige*; Nez (/ne/; nose) *ney/nex*; Drôle (/drol/; funny) *droal/droil*; Trop (/tro/; too much) *trow/trog*; Qui (/ki/; who) *kea/koa*; Douche (/duʃ/; shower) *dush/duth*; Dos (/do/; back) *dow/doy*; Faux (/fo/; false) *faw/faz*; Gris (/gri/; grey) *grea/grun*; Fleur (/flɛr/; flower) *flur/fler*.

Appendix E.

Filler pseudowords used in Experiments 1 to 3

chost; dath: dross; gide; gour; groat; gund; lain; largh; mest; nair; pard; pext; preen; queel;
shire; suil; warl; whigh; whure.

Appendix F.

Follow- up analyses on Experiments 1 and 2 considering orthographic similarity between experimental pseudowords and their corresponding base words.

In the first analysis, we conducted same model comparisons than in Experiments 1 and 2, except that we added to all compared models the continuous variable referred to as orthographic similarity (OS for each item). This follow- up analysis yielded the exact same patterns of results than described in the main text. In Experiment 1, Model 1 (AIC = 1059) had a better fit than model 0 (AIC = 1077, $X^2(1) = 20.01$, $p < .001$). The addition of the fixed effect of group to model 1 (model 2) did not improve adjustment compared to model 1 (AIC = 1061; $X^2 < 1$, n.s.). The model 3 (AIC = 1060, $X^2(1) = 2.59$, $p = .11$) did not improve the adjustment either. In Experiment 2, Model 1 (AIC = 1607) had a better fit than model 0 (AIC = 1645, $X^2(1) = 40.53$, $p < .001$). The addition of the fixed effect of group to model 1 (model 2) improved adjustment compared to model 1 (AIC = 1584; $X^2(1) = 25.15$, $p < .001$). The model 3 (AIC = 1585, $X^2 < 1$, n.s.) did not improve the adjustment.

In the second analysis, we compared the best fitted model from the first analysis with a model that contained a fixed effect of interaction between pseudoword status and OS (model 4). In Experiment 1, we found that adding this fixed effect of interaction improved the model adjustment compared to model 1 (AIC of model 4 = 1057, $X^2(1) = 4.73$, $p < .03$). And yet, visual examination of this interaction revealed that this effect was driven by the two items with lowest orthographic similarity indexes (i.e. quipe and nide from the base words keep and need) that showed stronger PsH effects, a result that goes against the predicted hypothesis and that should thus be taken with great caution. In Experiment 2, we found that model 4 that included the same random and fixed effects but also included a fixed effect of interaction between pseudoword status and OS did not provide a better adjustment than model 2 (AIC =

1584) that included random variables, a fixed effect of pseudoword status, of grade, and of OS (AIC of model 4 = 1583, $X^2(1) = 2.54$, $p = .11$, n.s.).

Acknowledgements

We wish to thank the head teachers and teachers of the French secondary schools who allowed the research to be carried out, and adolescents and their families for their cooperation in the project.

Funding

This study was supported by the grant n° W14RAT34 awarded from the University of Strasbourg and Mission Investissement d'Avenir (IdEx grant, 2014- 2015) to first author.