



This paper was originally published by the Cambridge University Press as:

Schröter, P., & Schroeder, S. (2018). **Differences in visual word recognition between L1 and L2 speakers: The impact of length, frequency, and orthographic neighborhood size in German children.** *Studies in Second Language Acquisition*, 40(2), 319–339.
<https://doi.org/10.1017/S0272263117000201>

This publication is with permission of the rights owner freely accessible due to an Alliance licence and a national licence (funded by the DFG, German Research Foundation) respectively.

Nutzungsbedingungen:

Dieser Text wird unter einer Deposit-Lizenz (Keine Weiterverbreitung - keine Bearbeitung) zur Verfügung gestellt. Gewährt wird ein nicht exklusives, nicht übertragbares, persönliches und beschränktes Recht auf Nutzung dieses Dokuments. Dieses Dokument ist ausschließlich für den persönlichen, nicht-kommerziellen Gebrauch bestimmt. Auf sämtlichen Kopien dieses Dokuments müssen alle Urheberrechtshinweise und sonstigen Hinweise auf gesetzlichen Schutz beibehalten werden. Sie dürfen dieses Dokument nicht in irgendeiner Weise abändern, noch dürfen Sie dieses Dokument für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, aufführen, vertreiben oder anderweitig nutzen. Mit der Verwendung dieses Dokuments erkennen Sie die Nutzungsbedingungen an.

Terms of use:

This document is made available under Deposit Licence (No Redistribution - no modifications). We grant a non-exclusive, nontransferable, individual and limited right to using this document. This document is solely intended for your personal, non-commercial use. All of the copies of this documents must retain all copyright information and other information regarding legal protection. You are not allowed to alter this document in any way, to copy it for public or commercial purposes, to exhibit the document in public, to perform, distribute or otherwise use the document in public. By using this particular document, you accept the above-stated conditions of use.

Provided by:

Max Planck Institute for Human Development
Library and Research Information
library@mpib-berlin.mpg.de

Research Article

DIFFERENCES IN VISUAL WORD RECOGNITION BETWEEN L1 AND L2 SPEAKERS

THE IMPACT OF LENGTH, FREQUENCY, AND ORTHOGRAPHIC NEIGH-
BORHOOD SIZE IN GERMAN CHILDREN

Pauline Schröter*

Max Planck Institute for Human Development, Berlin

Sascha Schroeder

Max Planck Institute for Human Development, Berlin

Abstract

Investigating the impact of linguistic characteristics on visual word recognition in children, we studied whether differences in native (L1) and second language (L2) processing already emerge at the beginning of reading development. German elementary school students in grades 2 to 6 completed a battery of standardized tests and a lexical decision task (LDT). Though L1 speakers outperformed L2 speakers on German skills, groups did not differ in their overall performance on the LDT. However, results from mixed-effect models revealed greater effects for word frequency and length in L2 over L1 speakers, indicating qualitative differences in the sensitivity to linguistic information between groups. This distinction persisted across all grades and after controlling for differences in vocabulary size and reading fluency. Findings extend evidence provided for adult L2 processing, suggesting that varying language exposure shapes the development of the word-recognition system already in the early stages of reading development.

It is widely accepted that lexical processing differs significantly between native (L1) and second language (L2) reading. This has been demonstrated within bilinguals by comparing visual word recognition in their L1 and L2 (e.g., de Groot, Borgwaldt, Bos, & van den Eijnden, 2002) as well as by comparing native speakers and bilingual L2 speakers in their word processing of the same language (e.g., Lemhöfer et al., 2008). Though evidence points to greater costs for word processing in L2 over L1, little is known about the differences between L1 and L2 speakers regarding the general mechanisms

*Correspondence concerning this article should be addressed to Pauline Schröter, Institute for Educational Quality Improvement, Humboldt-Universität zu Berlin, Unter den Linden 6, 10099 Berlin, Germany. E-mail: pauline.schroeter@iqb.hu-berlin.de

underlying the visual word-recognition system. These mechanisms are commonly investigated through within-language factors such as word frequency, length, or orthographic neighborhood size, whose marker effects present key issues in the cognitive modeling of reading. Several studies have provided evidence that L2 speakers have more difficulty recognizing low-frequency words than L1 speakers, which suggests that the impact of frequency is greater in L2 than in L1 processing. Yet, bilingual participants in these studies were adults who had acquired initial reading skills in their L1 at the same time as native speakers, and later learned how to read in their L2. The question that follows is whether it is varying exposure to print or differences in the mechanisms of lexical access that causes dissimilarities between L1 and L2 processing. An answer to that question would not only shed more light on the source of the frequency effect (FE), but also provide valuable insight into the development of the bilingual word-recognition system. The present article aims to address this issue by investigating visual word recognition in monolingual (L1) and bilingual (L2) German-speaking children at the beginning of initial reading acquisition. While controlling for factors associated with language exposure, the goal is to portray orthographic processing differences between L1 and L2 speakers regarding the impact of word frequency, length, and orthographic neighborhood size.

VISUAL WORD RECOGNITION IN BILINGUALS

Given the wealth of studies on bilingual language processing, it is of crucial importance to distinguish between different forms of bilingualism, whose characteristics have been shown to have different effects on lexical access. As opposed to balanced bilinguals, who are equally highly proficient in both of their languages, unbalanced bilinguals, who are usually less proficient in their L2 than in their L1, have been found to lag behind monolinguals on tasks involving lexical access. The costs reported in L2 speakers over L1 speakers include slower reaction times (RTs) in lexical decision tasks (LDTs; Brysbaert, Lagrou, & Stevens, 2016; Ransdell & Fischler, 1987), slower RTs and lower accuracy scores on picture naming (Gollan, Fennema-Notestine, Montoya, & Morris, 2005; Gollan, Montoya, Cera, & Sandoval, 2008; Ivanova & Costa, 2008) and category fluency tasks (e.g., Gollan, Montoya, & Werner, 2002; Rosselli et al., 2000), more tip of the tongue experiences (e.g., Gollan & Acenas, 2004; Gollan & Silverberg, 2001), and poorer word-identification skills through noise (Rogers, Lister, Febo, Basing, & Abrams, 2006). Taken together, there is reason to assume that speaking two languages has an impact on the development of visual word recognition.

To investigate differences in orthographic processing between L1 and L2, the mechanisms of lexical access need to be examined in detail. Yet, studies exploring the influence of linguistic characteristics on L2 word processing were exclusively carried out on adults, who had started to learn English on average at the age of 12 years. In 2002, de Groot, Borgwaldt, Bos, and van den Eijnden conducted a study with Dutch-English bilinguals, who completed a Dutch (L1) as well as an English (L2) LDT. The authors found that in both languages frequency variables were the most important predictors, while in English lexical decision was also affected by length. In 2008, Lemhöfer and colleagues compared native speakers of English to three different bilingual populations on an English progressive demasking task and found significant differences in the impact

of frequency-related measures between L1 and L2 speakers, which occurred irrespective of bilingual participants' L1. Transferring these findings to beginning readers, we expect the FE to be even more pronounced. Assuming the links between phonological, semantic, and orthographic representations, which are believed to develop as function of experience with print (Perfetti, 2007), are weaker in children than in adults, their lexical access should be more affected by differences in language exposure. Also, when learning to read in an orthographically transparent language like German as opposed to opaque English, effects of word length and neighborhood size should be more salient.

Because of their overall lack of orthographic representations at the very beginning of reading development, initially child L2 speakers should not differ from their L1 speaking peers in orthographic processing. Yet, considering that groups differ in their exposure to the target language, it is likely that differences will emerge during reading development. Though groups are equally immersed and exposed to formal reading instruction at school, less oral usage combined with less exposure to print in the target language at home should lead to a disadvantage in orthographic processing in L2 over L1 speakers. In the following, we will elaborate on the three marker effects in visual word-recognition research, hypothesizing potential differences between L1 and L2 speaking children.

WORD-FREQUENCY EFFECTS

There is consensus that the most robust predictor for language performance is the frequency of occurrence on a word in a language (for a review, see Brysbaert et al., 2011). The FE, which entails that high-frequency words are processed faster and more accurately than low-frequency words, is one of the most investigated phenomena in psycholinguistic research (Cop, Keuleers, Drieghe, & Duyck, 2015). Interactive activation models (e.g., Dijkstra & Van Heuven, 2002; McClelland & Rumelhart, 1981) suggest implicit learning to be the source of this effect (for an alternative account, see Kinoshita, 2015). Accordingly, repeated exposure to a lexical item raises this item's baseline activation, so that the activation threshold is reached earlier and lexical access to that item is executed faster (e.g., Monsell, 1991). However, because the maximum speed of lexical access processes is limited, the effect of facilitation saturates once an item has exceeded a certain amount of exposure (Morton, 1969). It is further assumed that lexical entries are usage based and that reduced exposure to a language leads to reduced lexical entrenchment, which describes the overall quality of lexical representations in the mental lexicon (Perfetti, 2007). Based on the assumption that L2 speakers' L2 representations have accumulated less exposure, the weaker links theory was proposed (Gollan, Montoya, Sera, & Sandoval, 2008), which posits the idea that over time reduced language practice leads to weaker links between orthography, phonology, and semantics. This causes greater processing costs in L2 compared to L1 especially for words in the lower-frequency range, which are naturally encountered less often and thus are even more affected by reduced exposure.

Several studies have provided evidence for a larger FE in L2 over L1 speakers (e.g., Van Wijnendaele & Brysbaert, 2002), with the strongest effect in bilinguals' L2, followed by their L1, and the smallest effect in monolinguals (Gollan et al., 2008). In a study comparing word identification times of monolingual adults to three different bilingual populations, Diependaele, Lemhöfer, and Brysbaert (2013) could show that the

magnitude of the FE depended on participants' vocabulary size and occurred irrespective of their bilingualism, with weaker effects for those with larger vocabularies. Kuperman and Van Dyke (2013) extended this finding by suggesting that the FE is a result of the impact of individual language exposure and vocabulary size on the accuracy of corpus-based frequency measures. Though these findings are based on data of adults with approximately 12 years of exposure to L2, it is likely that already at the beginning of reading development differences in the FE reflect differences in lexical entrenchment. According to the idea of statistical learning, which understands learning as a function of input, effects of repeated exposure to a word should be especially visible after the word was first acquired. This, in turn, suggests that the gap between the FE in L1 and L2 should be even more pronounced in children than in adults. For the present study, therefore, we hypothesize a larger FE in L2 than in L1 speakers.

WORD-LENGTH EFFECTS

A vast body of research has shown that during reading development the sensitivity to the length of a word decreases. The length effect (LE), which entails that with an increasing number of letters processing costs for a word become greater, has been taken as the main marker effect for the conversion of orthography to phonology. Designed to explain this conversion, the Dual-Route Cascaded model (DRC; Coltheart, Rastle, Perry, Langdon, & Ziegler, 2001) distinguishes between a sublexical, letter-by-letter decoding, strategy and a lexical strategy, in which all letters are processed in parallel. Within this framework, the decrease of the LE is often interpreted as a gradual transition from predominantly sublexical to relatively more lexical reading strategies, which goes along with increasing reading experience (e.g., Marinus, Nation, & de Jong, 2015). Already familiar words are stored in the mental lexicon and can be accessed as a whole, whereas unfamiliar words still need to be decoded letter by letter by means of grapheme phoneme conversion (GPC) rules. In other words, the larger the size of the lexicon, the less likely it is for a reader to come across an unfamiliar word and thus to use sublexical reading strategies. As a result, differences in vocabulary size affect the impact of word length on reading.

Within research on bilingualism, it is widely accepted that individuals who speak two languages have smaller vocabularies in each of their languages than monolingual speakers of either language (e.g., Bialystok, Luk, Peets, & Yang, 2010). Presumably, lexicalized concepts are distributed across a bilingual's two languages, such that some words are known in one language, some in the other, and only some in both (Oller, 2005). In L2 speakers, who often use their L2 only in a limited context, such as in school or for work, this imbalance is even more pronounced. In a study with Dutch children, Verhoeven and Vermeer (2005) found that at the end of elementary school L2 speakers' vocabulary was one quarter to one third smaller than that of their L1 speaking peers. This finding, commonly known as "the vocabulary gap" in research on L2 acquisition (Thordardottir, 2011), is mainly attributed to the fact that they divide their time between two language environments (e.g., Oller, Pearson, & Cobo-Lewis, 2007). There is ample evidence that the amount of language input strongly influences the degree of general language growth in both L1 and L2 speakers (e.g., Pearson, 2007). Due to their reduced exposure, during reading development L2 speakers are likely to encounter words that are

already familiar to L1 speakers but still unfamiliar to them. Consequently, they must rely on sublexical reading strategies while their native-speaking peers benefit from their greater number of lexical entries by using lexical strategies. Differences in the use of strategies should become more apparent the more the number of letters in a word increases, which is why we hypothesize a larger LE in L2 over L1 speakers. Especially in orthographically transparent languages, with consistent GPC rules like German, this should be most salient in participants' RTs.

ORTHOGRAPHIC NEIGHBORHOOD SIZE EFFECTS

Conceptualizing the development of visual word recognition as the growing ability to discriminate between orthographic patterns, orthographic neighborhood size, which is a broad metric of the similarity of a word to other words, also plays an important role. Yet, findings on the impact of the discriminability of a word, that is whether there are a great deal of words that look orthographically similar, are inconsistent, ranging from inhibitory effects (Grainger, 1990) over facilitatory effects (e.g., Andrews, 1989, 1992, 1997) to null effects (e.g., Keuleers, Lacey, Rastle, & Brysbaert, 2011; Lemhöfer et al., 2008). Several studies have demonstrated that neighborhood size affects word-recognition performance depending on the type of orthographic overlap. Exploring the effect of neighborhood size (NE) on form priming, Castles, Davis, Cavalot, and Forster (2007) found that while developing readers in grade 3 showed substantial priming effects for stimuli of two types of lexical similarity, two years later they only showed effects for stimuli of one type. In skilled adult readers, the same set of stimuli did not induce any priming effects. The authors interpreted this finding in terms of a tuning of the automatic word-recognition system, which develops as a function of vocabulary size. Early in reading acquisition, many of the similar-looking competitors of a target word are not yet familiar to the reader, which is why the system can afford to be tuned broadly and to also accept orthographically similar words as candidates for the target word. With growing vocabulary, however, the system must adapt to the presence of more competing words in the lexicon and thus employ a more finely tuned discrimination mechanism to recognize the target word. Based on their findings, Castles and colleagues postulated the lexical tuning hypothesis, which suggests that orthographic representations and thus the recognition system as such will become more precise during reading development. This view is consistent with findings provided by Andrews and Hersch (2010), who investigated the influence of individual differences on neighbor priming in adults. The authors found that while individuals with poor spelling skills showed facilitatory priming effects for high-neighborhood words, better spellers showed inhibitory effects. According to the lexical quality account (Perfetti, 1992), which argues that the precision of lexical representations is crucial for efficient orthographic processing, good spelling skills are an index of finely tuned lexical representations particularly for words with a higher amount of orthographic neighbors. Thus, based on the assumption that L2 speakers have weaker lexical representations, it could be hypothesized that they show a larger NE than their native-speaking peers. However, given L2 speakers' smaller vocabulary size and, consequently, their reduced knowledge of potential lexical competitors, a contrasting hypothesis would be that they are less affected by orthographic neighborhood size and thus show a smaller NE than L1 speakers. In the present study,

therefore, our analyses of differences in the effect of orthographic neighborhood size are of an exploratory rather than confirmatory nature.

THE PRESENT STUDY

Most research on L2 word processing is conducted with adults, that is, late L2 speakers who have prior reading experience in L1. To the best of our knowledge, no study has ever investigated the impact of within-language characteristics on L1 versus L2 visual word recognition in children at the beginning of reading development. To fill this gap, we examined effects of word frequency, length, and neighborhood size on single word reading in German elementary school students. To consider high intercorrelations between these variables, we used mixed-effect models for our analyses, which enabled us to simultaneously assess partial effects of several predictors while including random effects for items and participants. This way, we could estimate the impact of linguistic characteristics simultaneously and separately from the influence of the underlying sample and stimulus material. Our first goal was to find out whether differences reported for the FE in adult L1 versus L2 processing can already be observed in beginning readers. Second, we wanted to investigate whether, in contrast to adult L2 speakers, child L2 speakers would also differ from L1 speakers in the impact of length and neighborhood size. Third, we were interested in whether these differences were mediated by linguistic skills that are known to develop as a function of language exposure. For this reason, we assessed participants' vocabulary size and reading fluency. If potential differences in the impact of frequency, length, and neighborhood size between L1 and L2 speakers persisted after controlling for these factors, this would serve as evidence that processing differences between L1 and L2 speakers are not due to varying language exposure but to qualitative differences in the mechanisms of lexical access. Fourth, we aimed to study whether these differences would already show at the beginning of reading development or first emerge during elementary school. For this reason, we recruited children from different grades in elementary school, covering different stages of reading development and periods of exposure to print. In summary, with the present study we aimed to address the following four research questions:

- (1) Can differences reported for the FE in adult L1 versus L2 processing already be observed in beginning readers?
- (2) Do child L2 speakers also differ from L1 speakers in the impact of length and neighborhood size?
- (3) Are these differences mediated by vocabulary size or reading fluency?
- (4) Do differences in L1 versus L2 processing already show at the beginning of reading development or first emerge during elementary school?

METHOD

PARTICIPANTS

Within the frame of Developmental Lexicon Project, 621 children in grades 2 to 6 from seven elementary schools in Berlin participated in the study. Data collection was performed during regular school hours and comprised two sessions each lasting

45 minutes. In the first session, which was conducted in a group setting, children completed a battery of standardized tests as well as a questionnaire on their language background and social demographics. Participants who indicated to have bad vision or who had started to learn German later than at the age of 6 years were excluded from all analyses ($n = 10$). In the second session, which was a computerized single session, children completed an LDT comprising six blocks. Participants who showed high error rates ($> 50\%$), or performed $2.5 SD$ below their age mean in RT on the LDT were also excluded from all analyses ($n = 27$). Eventually, data from 189 second graders, 151 third graders, 127 fourth graders, and 117 sixth graders were analyzed. Children who reported to have never learned any other language but German were classified as L1 speakers, whereas children who reported to have a different native language were defined as L2 speakers. Ninety-nine percent of L1 speakers were born in Germany, while 1% reported to be born abroad and to have immigrated before the age of 6. Of L2 speakers, 86% were born in Germany and 14% reported to have immigrated before the age of 6. Altogether, L2 speakers spoke 26 different languages with various writing systems and orthographic depths. The most prevalent L1 was Turkish (29% of participants), followed by Arabic (13%), Russian (9%), English (8%), and Polish (7%). Sixty-six percent of L2 speakers reported to use both languages equally at home, while 17% indicated to mostly use German and 17% to mostly use their L1 at home. Though for the majority of them their L1 was the dominant language at the time of entering school, due to the growing exposure to L2 during elementary school German became their dominant language. Asked for their overall use of each language throughout the day, 60% reported to make more use of German, while 32% indicated to use both languages equally and 8% to make more use of their L1. As none of the participating schools offered language tuition for any other language but German and literary classes outside school are not very common in Germany, most L2 speakers could only speak in their L1 and learnt how to read and write only in German.

Overall sample characteristics, which were analyzed using analyses of variance (ANOVAs) with the between subject factor group (L1 vs. L2 speakers), are provided in Table 1. We assessed children's vocabulary size by administering the vocabulary subtest of the CFT-20R (Weiß, 2006), which is a multiple-choice power test that requires participants to select the closest-matching equivalent for a given target word. As expected, L2 speakers scored significantly lower than their monolingual peers. Reading fluency was measured using the Salzburger Lese-Screening 1–4 (grades 2–4; Mayringer & Wimmer, 2003) and 5–8 (grade 6; Auer, Gruber, Mayringer, & Wimmer, 2005), which are speed-reading tests that require participants to indicate whether sentences are true or false. Results indicated a significant advantage for L1 over L2 speakers on reading fluency. We also assessed nonverbal intelligence by administering the matrix subtest from the CFT-20R (Weiß, 2006), which is a power test that requires participants to complete a matrix by selecting the missing part. Here, results did not differ between groups.

STIMULUS MATERIAL

The stimulus set consisted of 1,152 German content words (768 nouns, 269 verbs, and 115 adjectives) and 1,152 pseudowords. *Word frequency* measures were taken from the childLex corpus (Schroeder, Würzner, Heister, Geyken, & Kliegl, 2015), which is based

TABLE 1. Means, standard deviations (in parentheses), and results from the ANOVAs of sample characteristics for L1 and L2 German speakers

	L1	L2	<i>F</i>	<i>df</i>	<i>p</i>
<i>N</i> (% female)	409 (51)	175 (55)			
Age (in years)	8.68 (1.49)	8.75 (1.60)	0.29	1, 582	.59
Vocabulary ^a	13.81 (6.82)	11.67 (6.77)	12.05	1, 580	<.001
Reading fluency ^b	35.77 (11.81)	32.09 (10.59)	12.62	1, 580	<.001
Nonverbal intelligence ^c	5.35 (2.36)	5.08 (2.55)	1.49	1, 580	.22

^aRaw scores: 0–30 (95% CI for L1 [13.14, 14.47] and L2 [10.65, 12.68])

^bRaw scores: 0–70 (95% CI for L1 [34.62, 36.92] and L2 [30.50, 33.67])

^cRaw scores: 0–12 (95% CI for L1 [5.12, 5.58] and L2 [4.70, 5.46])

on German children's literature and includes 10 million words. Measures refer to normalized type frequency, that is the number of occurrences of a distinct word form divided by the total number of words in the corpus, and ranged from 0.1 to 1044 ($M = 61.76$; $SD = 107.55$). We transformed frequency to log base 10, so that the size of the FE was not affected by a word's absolute occurrence in the corpus. *Word length* indicates the number of letters in each word, and ranged from 3 to 12 letters ($M = 6.0$; $SD = 1.81$). *Orthographic neighborhood size* was operationalized through the mean Levenshtein Distance from a word to its 20 closest orthographic neighbors (OLD20). We considered OLD20 to be the best measure for studying orthographic similarity effects because it enables a larger range of orthographic variability. In contrast to the standard *N* metric (Coltheart, Davelaar, Jonasson, & Besner, 1977), OLD20 is not restricted to letter transposition, but also allows for the deletion and insertion of letters (Yarkoni, Balota, & Yap, 2008). Consequently, orthographic neighbors do not have to have the same number of letters but can vary in length, which is especially important when measuring neighborhood size in a language with a great number of longer words like German. Measures were derived from the childLex corpus and ranged from 1 to 4.45 ($M = 1.72$; $SD = .57$).

Pseudowords were generated from target words using the multilingual pseudoword generator Wuggy (Keulers & Brysbaert, 2010), which is based on an algorithm that replaces subsyllabic elements (i.e., onset, nucleus, or coda) of words with equivalent elements from other words of the same language. All 1,152 pseudowords were pronounceable and matched their respective target word on length and capitalization (as in German nouns are always capitalized).

To collect responses for several stimuli from children of different age groups, we used a multimatrix design for the LDT. This means that each participant only worked on a small subset of the stimuli, which was adjusted in size to participants' grade (for a more detailed description of the multimatrix approach, see Schröter & Schroeder, 2017). To collect the same amount of responses for all stimuli for each grade, we first randomly assigned words to lists that differed in their number between grades. We used four lists with 288 words for grade 6, six lists with 192 words for grade 4, eight lists with 144 words for the beginning of grade 3 and six lists with 192 words for the end of grade 3, and eight lists with 144 words for grade 2. The same procedure was applied to pseudowords. Stimuli of each list were then randomly assigned to six blocks, of which each included a

different number of trials between grades, ranging from 96 trials for grade 6 to 64 for grade 4, 48 for the beginning of grade 3, 64 for the end of grade 3, and 48 trials for grade 2. The assignment of lists to participants was counterbalanced within each grade. The order of stimuli within each block was randomized for every participant.

PROCEDURE

The experimental software and testing apparatus were identical in each grade. Stimuli were presented on a 15-inch TFT monitor (60 Hz refresh rate, resolution 1028 × 768 pixels, placed at about 60 cm from the participants) on a Windows-compatible laptop (Intel Pentium, dual core 2.x GHz) running Inquisit 3.0. Manual responses were collected using the laptop's keyboard. Participants were instructed to decide whether a presented letter string formed a correct German word, and asked to press a green button on the keyboard for "yes" or a red button for "no" as quickly and accurately as possible. They also completed a practice block with four items and passed buffer items at the beginning of each new block. Each trial began with the presentation of a fixation cross for 500 ms in the center of the screen. After 500 ms, the target item appeared in the same place and remained on screen until the participant responded. There was an interstimulus interval of 500 ms after the response was given.

RESULTS

Accuracy scores and RTs were analyzed using mixed-effects models (Baayen, Davidson & Bates, 2008) as implemented in the *lme4* package (version 1.0-4; Bates, Maechler, Bolker & Walker, 2013) in the statistical software *R*. RT data were log-transformed and analyzed using a linear model, while accuracy data were logit-transformed and analyzed using a generalized linear model with a binomial link function. As our focus was on the processing of lexical characteristics, analyses of linguistic effects were performed on words only. In the main model (Model 1) Words and participants served as random effects, whereas group (L1 vs. L2 speakers) and its interactions with linguistic characteristics were included as fixed effects. Frequency, length, and neighborhood size were modeled as continuous predictors, centered, and included separately as linear, quadratic, and cubic effect components. Contrasts for post-hoc comparisons were estimated using the general linear hypotheses test generated with the *multcomp* package (Hothorn, Bretz, Westfall, Heiberger, Schuetzenmeister, 2014). To explore the role of linguistic skills as mediating factors, we fitted two additional models by separately adding vocabulary (Model 2) and reading fluency (Model 3) as fixed effects to the main model. To consider age-related differences in these skills, predictors were generated by z-transforming participants' raw scores of the tests on reading fluency and vocabulary for each grade. Both predictors were modeled continuously and included in the model as main effects together with their interactions with group and linguistic characteristics. To investigate whether effects emerge with increasing print exposure, we fitted a fourth model (Model 4) by adding grade (grade 2, 3, 4, and 6), its interactions with group and linguistic characteristics, and their three-way interactions as further fixed effects to the main model. For all models, we calculated marginal R^2 values, which represent the proportion of variance explained by the fixed factors alone, and conditional R^2 values, which describe the proportion of variance explained by both the fixed and random factors.

ACCURACY

See the first column of Table 2 for participants' mean accuracy scores across grades. For words, accuracy did not differ between groups ($t = 0.21, p = .42$). In their performance on pseudowords, L2 speakers were more error prone than their native-speaking peers ($t = 1.95, p = .03$). Post-hoc comparisons of groups within each grade, however, revealed that this effect was fully driven by group differences in grade 3 ($t = 2.94, p = .002$), which is why we did not interpret this finding as a general disadvantage for pseudoword rejection in L2 speakers. The first column of Table 3 presents the results of the main analysis. In the following, findings on the impact of frequency, length, and neighborhood size are described separately.

Word Frequency

We found a strong main effect of frequency as well as a significant interaction of frequency with group. As expected, the FE was larger in L2 than in L1 speakers. See Figure 1A for the shape of the FE in both groups. In L1 as well as in L2 speakers, the effect was nonlinear in nature. The interaction was driven by words in the lower-frequency range (-1 standard deviation from mean), which were responded to less accurately by L2 than by L1 speakers, $t = -3.9, p < .001$. For the higher-frequency range (+1 standard deviation), there was no processing difference between groups, $t = 1.22, p = .11$.

Word Length

There was no significant effect of length in any of the accuracy data. Given the nature of the effect, which typically emerges mainly in RT data, this was not surprising.

Neighborhood Size

Though we found a significant main effect for neighborhood size, indicating more errors for words with more neighbors, there was no difference in its impact between groups.

TABLE 2. Model means and standard errors (in parentheses) for RTs (in ms) and accuracy scores (in % correct) for L1 and L2 German speakers across grades

	Accuracy		Reaction Time	
	L1	L2	L1	L2
<i>Words</i>				
Grade 2	92 (0.6)	91 (1.0)	1969 (59)	1800 (73)
Grade 3	95 (0.4)	93 (0.9)	1315 (41)	1260 (66)
Grade 4	96 (0.3)	97 (0.6)	1029 (34)	1116 (72)
Grade 6	96 (0.4)	97 (0.4)	772 (30)	807 (39)
<i>Pseudowords</i>				
Grade 2	88 (1.0)	87 (1.5)	2932 (101)	2637 (123)
Grade 3	92 (0.7)	87 (1.9)	1916 (69)	1911 (115)
Grade 4	93 (0.7)	92 (1.6)	1486 (55)	1745 (130)
Grade 6	92 (0.9)	92 (1.1)	989 (44)	1090 (61)

Note: $N_{\text{grade 2}} = 189, N_{\text{grade 3}} = 151, N_{\text{grade 4}} = 127, N_{\text{grade 6}} = 117$.

TABLE 3. Main effects and interactions with group for accuracy and RTs from Model 1

	Accuracy			Reaction Time		
	χ^2	<i>df</i>	<i>p</i>	χ^2	<i>df</i>	<i>p</i>
<i>Fixed effects^a</i>						
Group	1.26	1	.26	0.03	1	.87
Frequency	426.02	3	<.001	505.50	3	<.001
Frequency \times Group	14.62	3	<.01	3.01	3	.39
Length	2.39	3	.49	442.08	3	<.001
Length \times Group	1.35	3	.72	36.47	3	<.001
Neighborhood	10.99	3	<.05	5.73	3	.13
Neighborhood \times Group	2.91	3	.41	1.91	3	.59
<i>Random effects^b</i>						
Words	2830			6365		
Participants	3758			108270		

Note: Tests are based on Type II sum of squares and χ^2 values with Kenward-Roger *dfs*.

^aMarginal R^2 values are .06 (Accuracy) and .05 (RT).

^bConditional R^2 values are .32 (Accuracy) and .75 (RT).

REACTION TIMES

For the RT analysis, incorrect trials and trials that deviated more than 2.5 *SD* from either the participant or item mean were discarded, accounting in sum for 8.3% of the raw data for words and 13.1% of the raw data for pseudowords. Participants' mean RTs across grades are shown in the second column of Table 2. Overall, there was no difference in RTs between L1 and L2 speakers, neither for words ($t = 0.05$, $p = .48$) nor for pseudowords ($t = 1.1$, $p = .14$). See the second column of Table 3 for the results of the main analysis. In the following, findings on the impact of word length, frequency, and neighborhood size are presented separately.

Word Frequency

As expected, we obtained a strong main effect of frequency. However, there was no significant interaction with group.

Word Length

As often found for visual word recognition in children, there was a strong LE in RTs, which also differed significantly between groups. As expected, planned post-hoc contrasts showed an overall greater LE for L2 over L1 speakers, indicating a processing advantage for L1 over L2 speakers with increasing word length. See Figure 1B for the shape of the LE in both groups. Whereas for words of shorter length (-1 standard deviation from mean) groups did not differ in their processing, $t < 0.03$, $p = .49$, L1 were faster than L2 speakers in their performance on words of longer length (+1 standard deviation), $t = -4.57$, $p < .001$. For L1 speakers, the progression of the LE was quadratic, meaning that the increase in RTs was steeper for words of shorter length and leveled out

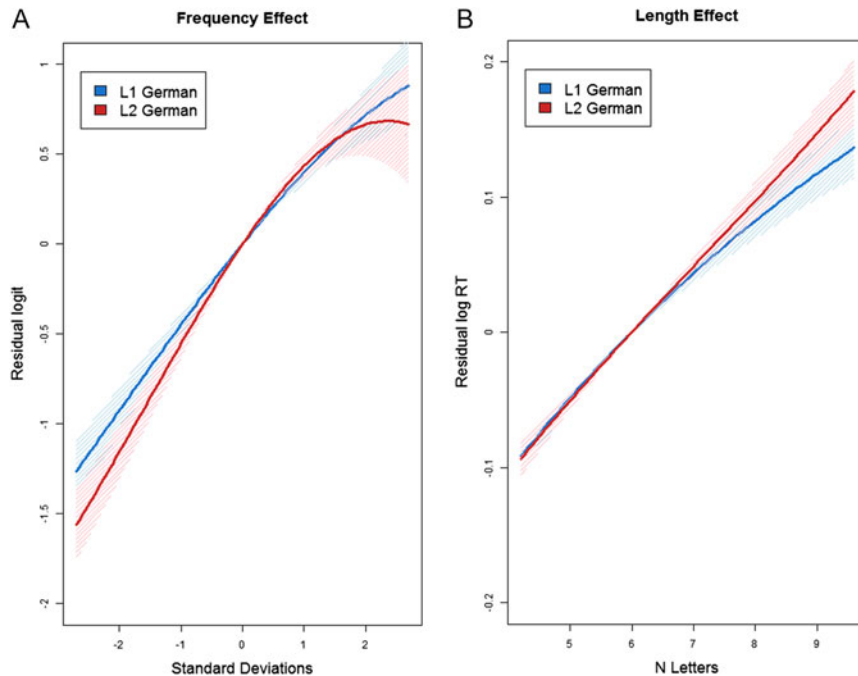


FIGURE 1. Differences in the shape of the FE (A) and LE (B) between L1 (blue line) and L2 (red line) German speaking children (with standard error polygons shaded in colors for groups).

for words of longer length. For L2 speakers, in contrast, the LE was exclusively linear in nature, characterized by a gradual increase in RT with a growing number of letters. More precisely, L1 speakers showed to be less susceptible to the impact of length after the number of six letters was reached, whereas L2 speakers were impacted gradually by every letter that was added.

Neighborhood Size

Though there was a tendency for words to be processed slower the more neighbors they had, we did not find a significant NE in any of the RT data.

ADDITIONAL ANALYSES

Vocabulary

Results of Model 2 are presented in Table 4. We found a reliable main effect of vocabulary in accuracy scores and RTs, suggesting better word-recognition skills with increasing vocabulary size. Results further showed significant interactions of vocabulary with length, frequency, and neighborhood size in both accuracy and RT data, indicating smaller effects for children with a larger vocabulary. Yet, analyses still revealed the Group \times Frequency interaction in accuracy scores as well as the Group \times Length

TABLE 4. Main effects and interactions with group and vocabulary from Model 2

	Accuracy			Reaction Time		
	χ^2	<i>df</i>	<i>p</i>	χ^2	<i>df</i>	<i>p</i>
<i>Fixed effects^a</i>						
Group	0.56	1	.45	5.50	1	<.05
Vocabulary	88.52	1	<.001	119.72	1	<.001
Frequency	427.93	3	<.001	514.07	3	<.001
Frequency \times Group	17.08	3	<.001	0.99	3	.80
Frequency \times Vocabulary	7.34	3	.37	66.67	3	<.001
Length	4.94	3	.18	420.09	3	<.001
Length \times Group	1.37	4	.71	18.99	3	<.001
Length \times Vocabulary	7.52	3	.06	255.16	3	<.001
Neighborhood	9.35	3	<.05	5.74	3	.13
Neighborhood \times Group	3.14	3	.37	6.01	3	.11
Neighborhood \times Vocabulary	18.48	3	<.001	53.27	3	<.001
<i>Random effects^b</i>						
Words	2922			6287		
Participants	2424			99982		

Note: Tests are based on Type II sum of squares and χ^2 values with Kenward-Roger *dfs*.

^aMarginal R^2 values are .08 (Accuracy) and .16 (RT).

^bConditional R^2 values are .32 (Accuracy) and .75 (RT).

interaction in RTs, which indicated that our findings on L1 and L2 speakers were independent of differences in their vocabulary size.

Reading Fluency

See Table 5 for the results of Model 3. Findings revealed a strong main effect for reading fluency in accuracy scores and RTs, suggesting more efficient word-processing skills with increasing reading fluency. In both accuracy and RT data, interactions of reading fluency with frequency, length, and neighborhood size reached significance, indicating smaller effects for children with higher reading fluency. Yet again, both the Group \times Frequency interaction in accuracy scores and the Group \times Length interaction in RTs were obtained, which represents further evidence that these effects occur irrespective of participants' reading fluency.

Grade

See Table 6 for the results of Model 4. As expected, there was a strong main effect of grade in accuracy scores as well as in RT data. Results for accuracy reproduced the Group \times Frequency interaction, but this effect did not interact with grade. Likewise, RT results produced the Group \times Length interaction, but did not reveal an additional effect of grade. We inferred from this that our results on the FE and LE between L1 and L2 speakers did not emerge over time, but persisted throughout the course of elementary school. While in accuracy scores there was only an interaction of grade with length, in the RT data we found interactions of grade with frequency, length, and neighborhood size,

TABLE 5. Main effects and interactions with group and reading fluency from Model 3

	Accuracy			Reaction Time		
	χ^2	<i>df</i>	<i>p</i>	χ^2	<i>df</i>	<i>p</i>
<i>Fixed effects^a</i>						
Group	0	1	.99	2.90	1	.09
Reading fluency	92.68	1	<.001	195.18	1	<.001
Frequency	434.86	3	<.001	513.93	3	<.001
Frequency \times Group	16.82	3	<.001	1.05	3	.79
Frequency \times Reading fluency	7.95	3	<.05	81.76	3	<.001
Length	5.07	3	.17	420.32	3	<.001
Length \times Group	1.84	3	.61	23.41	3	<.001
Length \times Reading fluency	9.54	3	<.05	334.37	3	<.001
Neighborhood	9.65	3	<.05	5.29	3	.15
Neighborhood \times Group	2.52	3	.47	3.08	3	.38
Neighborhood \times Reading fluency	8.40	3	<.05	39.11	3	<.001
<i>Random effects^b</i>						
Words	2892			6294		
Participants	2414			95877		

Note: Tests are based on Type II sum of squares and χ^2 values with Kenward-Roger *dfs*.

^aMarginal R^2 values are .09 (Accuracy) and .22 (RT).

^bConditional R^2 values are .32 (Accuracy) and .76 (RT).

suggesting a decline in the size of effects during reading development. The lack of significant interactions with group, however, indicated that the developmental pattern of L1 and L2 speakers was the same on all grade levels.

DISCUSSION

The goal of the present study was to investigate the impact of linguistic characteristics on visual word recognition in L1 and L2 German-speaking children. We investigated whether differences reported for the FE in adult L1 versus L2 processing can already be observed in beginning readers, and whether child L2 speakers also differ from child L1 speakers in the impact of length and neighborhood size. We additionally aimed to study whether these differences were mediated by vocabulary size or reading fluency, and whether they already show at the beginning of reading development or first emerge during elementary school. Within the frame of a cross-sectional mega study, L1 and L2 German elementary school children performed an LDT comprising words that varied in frequency, length, and orthographic neighborhood size. Despite their lower scores on vocabulary and reading fluency, results showed that quantitatively L2 speakers did not differ from L1 speakers in their word-recognition skills. Across all grades, there was no significant difference between the groups on overall accuracy scores or RTs, indicating that L2 speakers who start reading acquisition in their L2 quantitatively do not differ from native speakers on their visual word-recognition performance.

This is an important difference from findings on adult L2 speakers, who have acquired reading skills in L1 prior to L2 and whose L2 word-recognition skills have been shown to differ from native speakers' performance in the LDT (e.g., Brysbaert et al., 2016). That

TABLE 6. Main effects and interactions with group and grade from Model 4

	Accuracy			Reaction Time		
	χ^2	<i>df</i>	<i>p</i>	χ^2	<i>df</i>	<i>p</i>
<i>Fixed effects^a</i>						
Group	2.19	1	.14	0.40	1	.53
Grade	134.51	3	<.001	571.19	3	<.001
Group \times Grade	2.56	3	.46	5.15	3	.16
Frequency	398.71	3	<.001	510.53	3	<.001
Frequency \times Group	14.30	3	<.01	3.91	3	.27
Frequency \times Grade	11.33	3	.25	77.44	3	<.001
Frequency \times Group \times Grade	10.44	9	.32	11.42	9	.25
Length	2.43	3	.49	443.43	3	<.001
Length \times Group	1.55	3	.67	41.42	3	<.001
Length \times Grade	40.83	3	<.001	647.97	3	<.001
Length \times Group \times Grade	5.87	9	.75	15.87	9	.07
Neighborhood	10.16	3	<.05	5.86	3	.12
Neighborhood \times Group	2.25	3	.52	2.69	3	0.44
Neighborhood \times Grade	13.04	3	.16	30.10	3	<.001
Neighborhood \times Group \times Grade	9.99	9	.35	12.55	9	0.18
<i>Random effects^b</i>						
Words	2509			6341		
Participants	2413			65522		

Note: Tests are based on Type II sum of squares and χ^2 values with Kenward-Roger *dfs*.

^aMarginal R^2 values are .09 (Accuracy) and .39 (RT).

^bConditional R^2 values are .32 (Accuracy) and .76 (RT).

means that whether initial reading acquisition takes place in L1 or L2 seems to be crucial for the development of the word-recognition system. Even though they are initially learning how to read in their L2, child L2 speakers' word-recognition skills develop similarly to those of native speakers. Yet, based on the impact of within-language characteristics, our data revealed qualitative differences in the mechanisms of lexical access. In the following, we will address the first three of our four research questions by separately discussing our findings regarding the impact of frequency, length, and neighborhood size, respectively covering the mediating role of vocabulary size and reading fluency.

As expected, groups differed in size and shape of the FE, with a greater disadvantage for low-frequency words in L2 compared to L1 speakers. Especially in reading beginners, low-frequency words are most error prone, because they might not be known (yet) and therefore erroneously rejected as words in the LDT. Given that due to their smaller vocabulary size L2 speakers are more likely to treat them as pseudowords, it is not surprising that the Group \times Frequency interaction only showed in accuracy scores and not in RT data. Following studies on the FE in bilingual adults (e.g., Brysbaert et al., 2016; Cop et al., 2015; Diependaele, Lemhöfer, & Brysbaert, 2013), we explain this finding in terms of the lexical entrenchment account. Accordingly, due to their reduced exposure, L2 speakers' links between orthographic, phonological, and semantic representations are weaker, which in turn leads to greater processing costs especially for

words that are naturally encountered less often. However, most of these studies have been conducted with adults that had acquired their L2 at the end of childhood, that is who had roughly had 10 years of L2 exposure at the time of testing. Thus, our findings show that already a few years of reduced oral exposure seem to be sufficient to cause a greater effect of frequency. Given that the time of language exposure at school is the same for both groups, it seems that differences in the language environment alone suffice to impact children's subjective frequencies. Most interesting, however, is the finding that the Frequency \times Group interaction persisted after controlling for vocabulary size and reading fluency. This contrasts recent findings on the FE in adults, which show that once vocabulary size is considered, differences between L1 and L2 processing disappear (Brysbaert et al., 2016; Diependaele et al., 2013). Consequently, our results point to the presence of a qualitative distinction between lexical processing in L1 and L2 at the beginning of reading development. Differences in the impact of frequency between child L1 and L2 speakers seem to be imputed to a factor over and above vocabulary size and reading fluency, which poses a hypothesis that needs to be tested by further research.

Similarly, word length had a greater inhibitory effect in L2 than in L1 speakers, which was primarily driven by words of longer length. While L2 speakers showed to be gradually impacted by length, L1 speakers' sensitivity to length decreased after the number of six letters was reached. Within the framework of the DRC model, we interpreted this as a greater reliance on sublexical reading strategies in German L2 compared to L1 speakers. Especially for children in the early stages of reading development, many of the words are unfamiliar and thus treated like pseudowords. Presumably, children process a word sublexically until they reach a certain orthographic uniqueness point, at which the lexical representation is activated and the word will be recognized faster via the lexical route than the sublexical route. For L1 speakers, who rely on strong representations in their native language, decoding the first few letters of a word could be enough to trigger lexical activation. With both routes working in parallel, this will give less time for the sublexical route to influence the final output. L2 speakers, who have weaker lexical representations, are less likely to detect this point of uniqueness, which is why processing via the sublexical route continues. As this difference becomes more pronounced the more letters there are to decode, this would explain why for short words groups did not differ in their performance, while for longer words the LE is smaller in L1 than L2 speakers. Given that differences in lexical versus sublexical processing are expressed by longer RTs as a function of the increasing number of letters, it is thus not surprising that we found the LE only in RT data and not in accuracy scores. Yet, the fact that the Length \times Group interaction persisted after controlling for vocabulary size and reading fluency indicates again that an additional factor must be involved. We presume that given their smaller lexicon size L2 speakers have become accustomed to using sublexical reading strategies to such an extent that even after their lexicon has grown they overly rely on them. Alternatively, L2 speakers' grapheme phoneme mapping could lack in automaticity. Given that they have two phonological systems which they use every day, the activation process of language-specific phonemes might be slowed down, which in turn affects the mapping between graphemes and phonemes in a respective language. Both accounts would explain a later transition from sublexical to lexical reading strategies in L2 compared to L1 speakers, which is reflected by differences in the LE. Yet, given the scarcity of studies on orthographic processing in L2 reading

beginners, both accounts are highly hypothetical in nature and need further research to be supported.

With respect to the impact of orthographic neighborhood size, surprisingly, we did not find a difference between groups. This result is in line with Lemhöfer and colleagues' (2008) findings, who did not observe a difference between monolingual and bilingual speakers in the impact of various measures of orthographic neighborhood, either. Based on the inconsistency of findings from previous research on the NE, there are several possible explanations for the absence of this interaction in the present study. To begin with, the main effect of neighborhood size was rather small compared to the effects of frequency and length in our data. Given the overall little difference in accuracy scores between words of lower and higher OLD20, there was only a small likelihood for an interaction between neighborhood size and group to emerge at all. Also, most studies reporting a significant NE used a factorial designs and manipulated neighborhood size and word frequency at the same time, finding effects mainly for low-frequency words (e.g., Andrews, 1989). Mega studies, which include words from the entire continuum of linguistic characteristics, in contrast, often reported the lack of a significant NE (e.g., Keuleers et al., 2011). Discrepancies in results on the NE between mega studies and manipulation studies, therefore, might also arise from modeling orthographic neighborhood size as a continuous predictor versus including only extreme values. On a related note, differences in design could lead to a greater publication bias for research on the NE among studies using factorial designs compared to mega studies. In addition, the correlation of neighborhood size and word length in the present study was very high. Given that by using mixed-effect models we assessed partial effects of length and neighborhood size at the same time, we believe that in the present study effects of neighborhood size could be effects of length in disguise. In orthographically transparent languages, like German, readers have been found to rely more on smaller units (graphemes, phonemes), whereas in opaque languages, like English, larger units (bodies, rhymes) are more reliable (Ziegler, Perry, Jacobs, & Braun, 2001). Although a recent review of existing research on the grain size theory suggests that there are no differences in reliance on large units between German and English (Schmalz, Robidoux, Castles, Coltheart, & Marinus, 2017), there is evidence for differences in the reliance on small units (Rau, Moll, Snowling, & Landerl, 2015). This, in turn, could contribute to the difference in findings on the impact of neighborhood size in German compared to English. Especially in beginning readers of German, who do not yet have a large lexicon to rely on and thus depend on sublexical processing much more than skilled readers, word length seems to be of more relevance than neighborhood size. Yet, we need to emphasize that all the presented accounts are tentative explanations and need further research to be supported.

Turning to our fourth research question, results showed that all effects persisted after the impact of grade was considered. Although we expected group differences to emerge during reading development, there was no change in the pattern of effects throughout elementary school. While the start of initial reading instruction and the time of language exposure at school were the same for both groups, already in grade 2 differences in the impact of frequency and length were significant and continued to remain stable over time. This finding adds to the assumption that differences in the sensitivity to linguistic information between groups must be ascribed to a factor beyond exposure to print. This,

in turn, provides evidence that lexical access in child L2 speakers is genuinely different from L2 speaking adults, who have been shown to perform like native speakers if language exposure is controlled for (e.g., Diependale et al., 2013). Analyzing adult lexical decision data by using a diffusion model, Brysbaert, Lagrou, and Stevens (2016) recently found that even after vocabulary size was filtered out similar RTs in L1 and L2 speakers were not achieved in the same way. L2 speakers' mean drift rate, which is the model parameter describing the speed with which information accumulates until a certain threshold is exceeded, was lower than in L1 speakers. Though the authors stated that effects were not strong enough to refute the lexical entrenchment hypothesis, they explained this finding by suggesting that lexical information might build up more slowly in L2 than in L1 speakers. Acknowledging the fact that this is open for debate, we would like to adapt this view as a possible explanation for our results. Yet, this raises the question why, on a global level, groups do not differ in their word-recognition skills. Even though L2 speakers were more impacted by frequency and length than L1 speakers, and accordingly should have been disadvantaged on a task involving lexical access, they performed as fast and accurately on the LDT as their native-speaking peers. We believe, for that reason, that child L2 speakers might have a way to compensate for their greater susceptibility to within-language characteristics. In line with theories on the advantages of child bilingualism on reading acquisition, such as a better conceptual understanding of language (Bialystok, 2001), we suggest that their word-recognition system might be more flexible in nature. Although they differ from L1 speakers in the degree of automaticity when processing word frequency and length information, L2 speakers might, for instance, exceed their peers on processes inherent to lexical decision that we did not focus on in the present study. Yet, especially regarding the prevalent publication bias for bilingual advantages in cognitive processing (de Bruin, Treccani, & Della Sala, 2014), further research is needed to investigate which measures this could possibly be and whether a way of compensation like this can also be observed in other populations.

In summary, results of the present study show that though bilingual children learning to read in their L2 develop a word-recognition system similar to that of monolingual children, they show significant differences in their sensitivity to linguistic information. In contrast to L2 speakers with prior reading experience in L1, children who start initial reading acquisition in their L2 do not differ from native speakers in their overall word-recognition skills. We could observe the same difference reported for the FE in adult L1 versus L2 processing and further show that child L2 readers are more sensitive to word-length information than their native-speaking peers. Furthermore, our data provide evidence that in beginning readers differences in the impact of within-language characteristics between L1 and L2 are not mediated by vocabulary size or reading fluency but seem to be due to qualitative differences in the mechanisms of lexical access. Future research needs to investigate what these factors are and why their impact is confined to individuals who start initial reading acquisition in their L2. It would also be of value to study whether bilingual children who simultaneously learn to read in both of their languages show the same effects. For the time being, the data we have provided give further insight into the development of word-recognition processes and help to better understand reading acquisition in bilinguals. Especially given the increasing number of bilingual children who start schooling in their nonnative language, knowledge as such is essential to foster child L2 speakers' reading skills.

REFERENCES

- Andrews, S. (1989). Frequency and neighborhood effects on lexical access: Activation or search? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *15*, 802–814.
- Andrews, S. (1992). Frequency and neighborhood effects on lexical access: Lexical similarity or orthographic redundancy? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *18*, 234–254.
- Andrews, S. (1997). The effect of orthographic similarity on lexical retrieval: Resolving neighborhood conflicts. *Psychonomic Bulletin and Review*, *4*, 439–461.
- Andrews, S., & Hersch, J. (2010). Lexical precision in skilled readers: Individual differences in masked neighbor priming. *Journal of Experimental Psychology: General*, *139*, 299–318.
- Auer, M., Gruber, G., Mayringer, H., & Wimmer, H. (2005). *Salzburger Lese-Screening für die Klassenstufen 5–8 (SLS 5–8)*. Bern, Switzerland: Verlag Hans Huber.
- 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*, 390–412.
- Bates, D., Maechler, M., Bolker, B., & Walker, S. (2013). *lme4: Linear mixed-effects models using “Eigen” and S4. R package (Versions 1.0–4)* (<https://cran.r-project.org/web/packages/lme4/index.html>).
- Bialystok, E. (2001). *Bilingualism in development: Language, literacy, and cognition*. Cambridge: Cambridge University Press.
- Bialystok, E., Luk, G., Peets, K. F., & Yang, S. (2010). Receptive vocabulary differences in monolingual and bilingual children. *Bilingualism: Language and Cognition*, *13*, 525–531.
- Brysbaert, M., Lagrou, E., & Stevens, M. (2016). Visual word recognition in a second language: A test of the lexical entrenchment hypothesis with lexical decision times. *Bilingualism: Language and Cognition*, *1*, 1–19.
- Brysbaert, M., Buchmeier, M., Conrad, M., Jacobs, A. M., Bölte, J., & Böhl, A. (2011). The word frequency effect. *Experimental Psychology*, *58*, 412–424.
- Castles, A., Davis, C., Cavalot, P., & Forster, K. (2007). Tracking the acquisition of orthographic skills in developing readers: Masked priming effects. *Journal of Experimental Child Psychology*, *97*, 165–182.
- Coltheart, M., Davelaar, E., Jonasson, T., & Besner, D. (1977). Access to the internal lexicon. In S. Dornic (Ed.), *Attention and performance IV* (pp. 535–555). Hillsdale, NJ: Erlbaum.
- Coltheart, M., Rastle, K., Perry, C., Langdon, R., & Ziegler, J. (2001). DRC: A dual route cascaded model of visual word recognition and reading aloud. *Psychological Review*, *108*, 204–256.
- Cop, U., Keuleers, E., Drieghe, D., & Duyck, W. (2015). Frequency effects in monolingual and bilingual natural reading. *Psychonomic Bulletin and Review*, *22*, 1216–1234.
- De Bruin, A., Treccani, B., & Della Sala, S. (2014). Cognitive advantage in bilingualism: An example of publication bias? *Psychological Science*, *26*, 99–107.
- De Groot, A. M. B., Borgwaldt, S., Bos, M., & van den Eijnden, E. (2002). Lexical decision and word naming in bilinguals: Language effects and task effects. *Journal of Memory and Language*, *47*, 91–124.
- Diependaele, K., Lemhöfer, K., & Brysbaert, M. (2013). The word frequency effect in first- and second-language word recognition: A lexical entrenchment account. *The Quarterly Journal of Experimental Psychology*, *66*, 843–863.
- Dijkstra, T., & Van Heuven, W. J. (2002). The architecture of the bilingual word recognition system: From identification to decision. *Bilingualism: Language and Cognition*, *5*, 175–197.
- Gollan, T. H., & Acenas, L.-A. R. (2004). What is a TOT? Cognate and translation effects on tip-of-the-tongue states in Spanish–English and Tagalog–English bilinguals. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *30*, 246–269.
- Gollan, T. H., & Silverberg, N. B. (2001). Tip-of-the-tongue states in Hebrew–English bilinguals. *Bilingualism: Language and Cognition*, *4*, 63–83.
- Gollan, T. H., Montoya, R. I., & Werner, G. A. (2002). Semantic and letter fluency in Spanish–English bilinguals. *Neuropsychology*, *16*, 562–576.
- Gollan, T. H., Montoya, R. I., Sera, C., & Sandoval, T. C. (2008). More use almost always means a smaller frequency effect: Aging, bilingualism, and the weaker links hypothesis. *Journal of Memory and Language*, *58*, 787–814.
- Gollan, T. H., Montoya, R. I., Fennema-Notestine, C., & Morris, S. K. (2005). Bilingualism affects picture naming but not picture classification. *Memory and Cognition*, *33*, 1220–1234.

- Grainger, J. (1990). Word frequency and neighborhood frequency effects in lexical decision and naming. *Journal of Memory and Language*, *29*, 228–244.
- Hothorn, T., Bretz, F., Westfall, P., Heiberger, R. M., & Schuetzenmeister, A. (2014). *Multcomp: Simultaneous inference in general parametric models. R package (Version 1.0–3)* (<https://cran.r-project.org/web/packages/multcomp/index.html/>).
- Ivanova, I., & Costa, A. (2008). Does bilingualism hamper lexical access in speech production? *Acta Psychologica*, *127*, 277–288.
- Keuleers, E., & Brysbaert, M. (2010). Wuggy: A multilingual pseudoword generator. *Behavior Research Methods*, *42*, 627–633.
- Keuleers, E., Lacey, P., Rastle, K., & Brysbaert, M. (2011). The British Lexicon Project: Lexical decision data for 28,730 monosyllabic and disyllabic English words. *Behavior Research Methods*, *44*, 287–304.
- Kinoshita, S. (2015). Visual word recognition in the Bayesian reader framework. In A. Pollatsek (Ed.), *The Oxford handbook of reading* (pp. 63–76). Oxford, UK: Oxford University Press.
- Kuperman, V., & Van Dyke, J. A. (2013). Reassessing word frequency as a determinant of word recognition for skilled and unskilled readers. *Journal of Experimental Psychology: Human Perception and Performance*, *39*, 802–823.
- Lemhöfer, K., Dijkstra, T., Schriefers, H., Baayen, R. H., Grainger, J., & Zwitserlood, P. (2008). Native language influences on word recognition in a second language: A megastudy. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *34*, 12.
- Marinus, E., Nation, K., & de Jong, P. F. (2015). Density and length in the neighborhood: Explaining cross-linguistic differences in learning to read in English and Dutch. *Journal of Experimental Child Psychology*, *139*, 127–147.
- Mayringer, H., & Wimmer, H. (2003). *Salzburger Lese-Screening für die Klassenstufen 1–4 (SLS 1–4)*. Bern, Switzerland: Verlag Hans Huber.
- McClelland, J. L., & Rumelhart, D. E. (1981). An interactive activation model of context effects in letter perception: I. An account of basic findings. *Psychological Review*, *88*, 375–407.
- Monseil, S. (1991). The nature and locus of word frequency effects in reading. In D. Besner & G. W. Humphreys (Eds.), *Basic processes in reading: Visual word recognition* (pp. 148–197). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Morton, J. (1969). Interactions of information in word recognition. *Psychological Review*, *76*, 165–178.
- Oller, D. K. (2005). The distributed characteristic in bilingual learning. In J. Cohen, K. T. McAlister, K. Rolstad, & J. MacSwan (Eds.), *ISB4: Proceedings of the 4th International Symposium on Bilingualism* (pp. 1744–1749). Somerville, MA: Cascadilla Press.
- Oller, D. K., Pearson, B. Z., & Cobo-Lewis, A. B. (2007). Profile effects in early bilingual language and literacy. *Applied Psycholinguistics*, *28*, 191–230.
- Pearson, B. Z. (2007). Social factors in childhood bilingualism in the United States. *Applied Psycholinguistics*, *28*, 399–410.
- Perfetti, C. A. (1992). The representation problem in reading acquisition. In P. B. Gough, L. C. Ehri, & R. Treiman (Eds.), *Reading acquisition* (pp. 145–174). Hillsdale, NJ: Erlbaum.
- Perfetti, C. A. (2007). Reading ability: Lexical quality to comprehension. *Scientific Studies of Reading*, *11*, 357–383.
- Ransdell, S. E., & Fischler, I. (1987). Memory in a monolingual mode: When are bilinguals at a disadvantage? *Journal of Memory and Language*, *26*, 392–405.
- Rau, A. K., Moll, K., Snowling, M. J., & Landerl, K. (2015). Effects of orthographic consistency on eye movement behavior: German and English children and adults process the same words differently. *Journal of Experimental Child Psychology*, *130*, 92–105.
- Rogers, C. L., Lister, J. J., Febo, D. M., Besing, J. M., & Abrams, H. B. (2006). Effects of bilingualism, noise, and reverberation on speech perception by listeners with normal hearing. *Applied Psycholinguistics*, *27*, 465–485.
- Rosselli, M., et al. (2000). Verbal fluency and repetition skills in healthy older Spanish-English bilinguals. *Applied Neuropsychology*, *7*, 17–24.
- Schmalz, X., Robidoux, S., Castles, A., Coltheart, M., & Marinus, E. (2017). German and English bodies: No evidence for cross-linguistic differences in preferred grain size. *Collabra: Psychology*, *3*, 5.
- Schroeder, S., Würzner, K.-M., Heister, J., Geyken, A., & Kliegl, R. (2015). ChildLex: A lexical database of German read by children. *Behavior Research Methods*, *47*, 1085–1094.

- Schröter, P., & Schroeder, S. (2017). The Developmental Lexicon Project: A behavioral database to investigate visual word recognition across the lifespan. *Behavior Research Methods*, 1–21. doi:10.3758/s13428-016-0851-9.
- Thordardottir, E. (2011). The relationship between bilingual exposure and vocabulary development. *International Journal of Bilingualism*, 15, 426–445.
- 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.
- Verhoeven, L., & Vermeer, A. (2005). Het Ongelijk van Netelenbos? Toetsing van Kleuters en hun Prestaties op de Cito Eindtoets Basisonderwijs. *Toegepaste Taalwetenschap in Artikelen*, 74, 123–133.
- Weiß, R. H. (2006). *Grundintelligenztest Skala 2 (CFT 20-R) mit Wortschatztest (WS) und Zahlenfolgentest (ZF)*. Hogrefe: Verlag für Psychologie.
- Yarkoni, T., Balota, D., & Yap, M. (2008). Moving beyond Coltheart's N: A new measure of orthographic similarity. *Psychonomic Bulletin and Review*, 15, 971–979.
- Ziegler, J. C., Perry, C., Jacobs, A. M., & Braun, M. (2001). Identical words are read differently in different languages. *Psychological Science*, 12, 379–384.