Bilingualism: Language and Cognition

cambridge.org/bil

Research Article

Cite this article: Parshina O, Laurinavichyute AK, Sekerina IA (2020). Eye-movement benchmarks in Heritage Language reading. *Bilingualism: Language and Cognition* 1–14. https://doi.org/10.1017/S136672892000019X

Received: 20 February 2019
Revised: 2 February 2020
Accepted: 12 February 2020

Keywords:
bilingualism; heritage language; reading; eye movements; Russian; children; L2 learners

Address for correspondence: Olga Parshina, E-mail: oparshina@gradcenter.cuny.edu

Eye-movement benchmarks in Heritage Language reading

Olga Parshina[1], Anna K. Laurinavichyute[2,3] and Irina A. Sekerina[1,2]

[1]College of Staten Island and The Graduate Center of The City University of New York; [2]National Research University Higher School of Economics, Russian Federation and [3]University of Potsdam, Germany

Abstract

This eye-tracking study establishes basic benchmarks of eye movements during reading in heritage language (HL) by Russian-speaking adults and adolescents of high (n = 21) and low proficiency (n = 27). Heritage speakers (HSs) read sentences in Cyrillic, and their eye movements were compared to those of Russian monolingual skilled adult readers, 8-year-old children and L2 learners. Reading patterns of HSs revealed longer mean fixation durations, lower skipping probabilities, and higher regressive saccade rates than in monolingual adults. High-proficient HSs were more similar to monolingual children, while low-proficient HSs performed on par with L2 learners. Low-proficient HSs differed from high-proficient HSs in exhibiting lower skipping probabilities, higher fixation counts, and larger frequency effects. Taken together, our findings are consistent with the *weaker links* account of bilingual language processing as well as the *divergent attainment* theory of HL.

1. Introduction

Heritage language (HL) has only recently become a valued source of data for research in theoretical linguistics due to its unique properties. Heritage speakers (HSs) are early bilinguals who were raised speaking the minority language but switched to the majority language in later childhood (Valdés, 2000). In many aspects of language production and comprehension, they differ from other bilinguals such as second language learners. Despite the early exposure to the HL, most HSs have very limited literacy skills, especially when the HL orthography is different from the majority language, as in the case of Arabic, Chinese, Korean, and Russian, with English being a majority language in the USA (Benmamoun, Montrul & Polinsky, 2013; Carreira & Kagan, 2011; Koda, Zhang & Yang, 2008; Polinsky, 2018; Polinsky & Kagan, 2007; Xiao, 2006). In this eye-tracking study, we investigate reading skills of bilingual adult and adolescent speakers of Russian, non-Roman-based HL, and explore a connection between reading and the proficiency level in HL.

We start with a brief description of universal eye-movement measures in reading and how they are affected by lexical properties of the words. Then we present the novel *Bilingual Russian Sentence Corpus (BiRSC)* that focuses on reading skills in two groups of bilingual heritage Russian-English HSs and compare them to the previously studied two comparison groups: namely, Russian-speaking monolingual skilled readers (RSC; Laurinavichyute, Sekerina, Alexeeva, Bagdasaryan & Kliegl, 2019) and Russian-speaking monolingual children learning to read (Korneev, Akhutina & Matveeva, 2017). We add another comparison group, L2 learners of Russian, and demonstrate that the proficiency in HL only weakly influences eye-movement characteristics in reading as HL reading patterns regardless of their proficiency, were more similar to L2 learners and children than monolingual adults, a finding consistent with current bilingual language processing models and theories of HL acquisition.

1.1. Eye-movement benchmarks in monolingual reading

The start of the basic eye-movement research in reading goes back to the early 20th century (Huey, 1908) and since then language researchers have been using eye-movement measures to test various psycholinguistic theories (Radach & Kennedy, 2004; Rayner, 2009) as well as the models of eye-movement control (Reichle, Pollatsek, Fisher & Rayner, 1998; Engbert, Nuthmann, Richter & Kliegl, 2005). These measures are traditionally classified either as EARLY or LATE (see Table 1 for the list of measures and their description), depending on what stage of language processing they reflect (Clifton, Staub & Rayner, 2007; Roberts & Siyanova-Chanturia, 2013). Early measures are generally sensitive to lexical access, early information integration, and early morphological decomposition (but see Vasishth, von der Malsburg & Engelmann, 2013, for discussion). Late measures are indicative of post-lexical processing wherein they reflect reanalysis and recovery from difficulties in morphosyntactic processing as well as semantic integration (Boston, Hale, Kliegl, Patil & Vasishth, 2008).

Table 1. Early and late eye-movement measures in reading.

Abbreviation	Measure
EARLY:	
	Initial saccade landing position: reflects pre-lexical processing and is responsible for some of the subsequent saccade decisions.
FFD	*First fixation duration*: the duration of the first fixation on a word
SFD	*Single fixation duration* for words that are fixated only once during first pass reading
GD	*Gaze duration*: the sum of all fixations on the word before the eyes move elsewhere
P0	*Probability* of skipping the word
P1	*Probability* of fixating the word only once
LATE:	
TT	*The sum of all fixation durations* on the word
R0	*Probability* of regression saccade to the previous word from the current word
RG	*Probability* of regressing back to the word from the following word
P2+	*Probability* of fixating the word more than once

Lexical properties of the word that influence eye movements, i.e., frequency, predictability, and length, are also well-established (for review, see Staub & Rayner, 2007). Typically, high-frequency words are recognized faster compared to the low-frequency words, the phenomenon known as *the word frequency effect* (FE, e.g., Inhoff & Rayner, 1986). In terms of eye-movement measures, the effect leads to shorter mean fixation durations and higher skipping probabilities on the high-frequency words. The same findings are observed for words with high predictability. Conversely, reading times increase as the length of the word increases whereas the skipping likelihood decreases (Rayner, 1998).

The effects of lexical properties on eye movements are universal across languages despite differences in grammar and script, supporting the theory of the "universal science of reading" (Share, 2008). The basic eye-movement characteristics, on the other hand, vary dependent on the writing script (e.g., alphabetic vs. logographic), visual and informational density (Liversedge, Drieghe, Li, Yan, Bai & Hyönä, 2016). In English, the mean fixation durations in monolingual adults typically range between 220–250 ms as a function of the word's length, frequency and predictability as well as its fit in semantic and syntactic context of the sentence. Approximately one-third of all words are skipped, with skipping probabilities contingent on the word length (i.e., 3-letter words are skipped 70% of the time, 7-letter ones only 20%; function words are skipped more often than content words). The probability of regressions is about 20%, but it varies depending on the syntactic and semantic complexity of the text.

The saccade landing position is also dependent on the writing system. This measure provides information about the preferred viewing position within the word and serves as the basis of comparison with THE OPTIMAL VIEWING POSITION (OVP). O'Regan and Jacobs (1992) demonstrated that in cases where the first eye fixation is in OVP, the word is recognized faster and requires fewer re-fixations. In alphabetic languages with Latin script, such as English, German or French, OVP is located closer to the center of the word (Hyönä & Bertram, 2011; Nuthmann, Engbert & Kliegl, 2005; Vitu, McConkie, Kerr & O'Regan, 2001). In logographic non-Roman-based languages, such as Chinese, it varies from the center of the word that was fixated only once to the beginning of the word that has multiple fixations (Li, Liu & Rayner, 2011). In Arabic, the OVP depends on the word length, wherein short words attract center-based saccade landing and long words receive the first saccade at the beginning of the word (Paterson, Almabruk, McGowan, White & Jordan, 2015).

In monolingual children (findings from English and Finnish), eye-movement benchmarks reflect their developmental nature, with longer mean fixation durations (280–300 ms), lower skipping rate (ranging from 9% to 39% of all words), and more regressions (30%). In comparison to adults, children are slower with grapheme-to-phoneme decoding process compared to adults due to less experience with written materials and smaller visual perceptual span. Similar to adults, children as young as 7 years old tend to land the first saccade closer to the word center (Barnes & Kim, 2016; Blythe & Joseph, 2011; Joseph, Liversedge, Blythe, White & Rayner, 2009; Vitu et al., 2001). The differences in reading skills gradually disappear with age (Blythe, Häikiö, Bertam, Liversedge & Hyönä, 2011; Häikiö, Bertram, Hyönä & Niemi, 2009; Mancheva, Reichle, Lemaire, Valdois, Ecalle & Guérin-Dugué, 2015).

1.2. Eye-movement benchmarks in bilingual L2 reading

Only recently have eye movements in reading become a more common object in bilingualism research. The primary goal of these studies is to use the eye-movement characteristics in L2 reading to tap into key debates of bilingual language processing, such as lexical access, organization of bilingual mental lexicon, and grammatical and discourse processing (for review, see Roberts & Siyanova-Chanturia, 2013).

Early eye-movement measures (see Table 1) reflect lexical access – therefore, they could be used to empirically test theories of word representations in the bilingual lexicon, such as THE WEAKER LINKS account (Gollan, Montoya, Cera & Sandoval, 2008), according to which the reduced exposure and less accumulated practice of both of the bilingual's languages lead to the weakened links between word forms and their mental representations. As a result, lexical access in non-dominant language is slowed down as reflected in larger frequency effects and longer reading times. Proficiency, however, should modulate these effects because high-proficient L2 bilinguals have higher amounts of non-dominant language exposure.

Indeed, in recent L2 corpus eye-movement studies with unbalanced L1 Dutch-L2 English readers, Cop and colleagues (Cop, Drieghe & Duyck; 2015; Cop, Dirix, Drieghe & Duyck, 2017) found longer total reading times (1523 ms), longer average fixation durations (239 ms), more fixation counts (8.3 per sentence), and decreased skipping probability (48% vs. 52% in L1) in reading in a non-dominant language. In that sense, L2 readers resemble young monolingual children who just started to learn to read (Blythe & Joseph, 2011; Blythe et al., 2011) or low-literate monolingual adults (Barnes & Kim, 2016; Kuperman & Van Dyke, 2011). However, the influence of L2 proficiency on sentence level reading parameters was small (only the fixation count was affected) suggesting that L2 reading can be close to L1 but quite variable. There was no difference in bilinguals' L1 and monolingual reading which implies that, for late bilinguals, the strength of links between lexical representations and word forms for L1 words is comparable to that of monolingual readers.

Whitford and Titone (2012, 2016, 2017) who recorded eye movements of late L2 bilinguals of various ages (18–86 years), different language dominance (English or French), and proficiency found that the amount of exposure to the weaker language determined the magnitude of the frequency effect, with lower levels of exposure leading to larger frequency effects in early and late eye-movement measures in L2. Berzak, Katz and Levy (2018) recently demonstrated that the proficiency level in L2 could affect eye movements even more directly. They found that first fixation duration and total reading times not only correlated with standardized tests of English language proficiency (*MET*: $r = .5$ and *TOEFL*: $r = .54$) but were also effective in predicting the outcomes of these tests.

To summarize: monolinguals and high-proficient bilingual adult readers do not show significant qualitative differences in eye-movement measures in L1 reading (cf. Whitford & Joanisse, 2018, for differences in bilingual children). However, eye movements in low-proficient L2 readers are characterized by longer fixation durations, lower likelihood of skipping the words, higher probabilities of regressive saccades, and larger frequency effects making them similar to monolingual children or monolingual adults with poor reading skills. Such empirical findings support the weaker links account (Gollan et al., 2008) that puts exposure to non-dominant language as the bottleneck factor affecting various aspects of non-dominant language processing, including acquisition and fluency of reading skills.

1.3. Reading in a heritage language

Little is known about literacy skills of young adult HSs whose HL development was interrupted by the start of school in the dominant language. With respect to reading, previous research suggests that there are no benefits of the early exposure to literacy in HL. Earlier studies that compared Chinese L2 learners and Chinese HSs reported no difference between these two groups in Chinese character recognition, reading comprehension or vocabulary knowledge (Ke, 1998; Xiao, 2006). More recent HL research revealed only the facilitatory effect of vocabulary on reading comprehension in HL (Zhang & Koda, 2018).

Literacy remains the weakest domain in HSs in comparison to L2 learners who outperform them on a variety of written tasks, even when the dominant and HL languages share the same script, such as English and Spanish (Keating, VanPatten & Jegerski, 2011; Potowski, Jegerski & Morgan-Short, 2009; Tse, 2001). While the reasons for L2 advantage in written tasks is a topic for separate research, one can think that the possible explanation lies in the formal setting of the L2 acquisition. HSs acquire language at home and, in most cases, have very limited experience with its written form, whereas L2 speakers receive formal classroom instruction aimed to speed up literacy acquisition.

In contrast to literacy, oral competency in HL is consistently linked to better performance of HSs in auditory tasks compared to L2 learners. For example, Gor, Cook, Pandza and Chrabaszcz (2018, February) compared the performance of 28 Russian HSs and 31 L2 learners in a grammaticality judgment task (GJT) in visual and auditory modalities. The results showed that, while there was no difference in accuracy scores among groups on the visual GJT, HSs outperformed proficiency-matched L2 learners in the auditory modality. Considerably better auditory language processing skills are also reported for Chinese HSs (Xiao, 2006) and Spanish HSs (Potowski et al., 2009). From a pedagogical perspective, it is important to identify how literacy skills differ between HSs and L2 learners so that they can be targeted for training in heritage or mixed language learning classrooms.

As of yet, eye-movement benchmarks in reading in HL remain unknown, and their investigation is the primary goal of the present study: This is the first eye-tracking study of reading in HL of non-Roman-based orthography that compares HSs to monolingual children learning to read and L2 learners and explores a connection between reading in HL and the proficiency level. We seek answers to research questions in (1)-(3):

(1) How does proficiency (high vs. low) in HL defined by simple reading tests affect eye movements in reading isolated sentences?
(2) Whose eye-movement patterns do HSs' patterns resemble the most: those of adults, L2 learners or monolingual Russian children learning to read?
(3) Will canonical effects of word length and frequency (both foveal and parafoveal) hold for reading in HL?

2. Eye-movement benchmarks in reading in Russian

The non-Roman-based Cyrillic alphabet, complex polysyllabic structure of the words, and morphological richness distinguish Russian from many other European languages (for overview of Russian orthography, see Rakhlin, Kornilov & Grigorenko, 2017) and make it an ideal target for comparative reading research in alphabetic languages. Despite the fact that this method of eye-movement recordings was tested in the Soviet Union as early as in the 1960s by Alfred Yarbus (1967), until recently there were no studies that investigated eye-movement benchmarks in Cyrillic in monolingual Russian adults or acquisition of reading in Russian children. To the best of our knowledge, our study that established *the Russian Sentence Corpus* (*RSC*, Laurinavichyute et al., 2019) was the first one to fill in this gap. Eye movements of monolingual Russian 8-year-old children were also only recently investigated, using a design parallel to the RSC (Korneev, Akhutina & Matveeva, 2017). In order to test our predictions regarding eye-movement benchmarks and literacy skills of HL Russian young adults, we will use the previously collected data from skilled monolingual readers and monolingual children and compare our novel data form HSs and L2 learners to them.

2.1. Russian Sentence Corpus (RSC) and its child version

The RSC design follows the cross-linguistic protocol of *the Potsdam Sentence Corpus* for German (Kliegl, Grabner, Rolfs & Engbert, 2004) and includes basic eye-movement characteristics of 96 skilled monolingual Russian readers (66 women, $M_{\text{Age}} = 24$, range 18–80). The RSC is based on 144 sentences randomly selected and modified from the Russian National Corpus that represent various types of grammatical structures typical of the Russian language. Each sentence contains a target word orthogonally manipulated in a $3 \times 3 \times 2$ design: the part-of-speech (adjectives, nouns, verbs), length (short, medium and long), and frequency (either high:> 50 ipm, or low: <10 ipm). The dependent measures for the RSC are listed in Table 1. Although there were some language-specific differences (see in Laurinavichyute et al., 2019 for discussion), in general the eye-movement benchmarks in reading in Russian presented in Table 4 (the first column) were consistent with findings for other alphabetic languages and

Table 2. Participant characteristics and average scores for performance on 4 reading assessment tasks.

	Group		
	High-proficient HSs Mean (*SD*)	Low-proficient HSs Mean (*SD*)	L2 Learners Mean (*SD*)
N	21	27	27
Age (y.o)	17.52 (3.4)	19.07 (3.68)	21.04 (6.99)
Gender (women:men)	13:8	22:5	17:10
Age of Arrival to USA (years)	5.10 (5.42)	2.33 (4.26)	0.15 (0.77)
Age of Reading start in Russian (years)	4.52 (2.11)	10.26 (5.75)	17.67 (5.84)
Daily Russian language exposure (%)	35.10 (17.31)	22.37 (17.21)	6.81 (5.8)
Daily reading exposure to Russian (min)	30–60	0–30	0–30
Self-reported proficiency measures in Russian (scale 1–5, with 5 the highest)			
Comprehension	4.28 (.783)	3.44 (.892)	2.44 (.801)
Speaking	4.24(.625)	3.41 (.971)	2.59 (.844)
Reading	4.00 (.837)	2.93 (.616)	2.63 (.926)
Writing	3.86 (.91)	2.56 (1.01)	2.74 (.944)
Reading objective assessments (scores)			
Word ID-Rus	44.67 (1.53)	38.33 (6.2)	35.11 (10.3)
Word ID-Eng	38.95 (4.0)	41.03 (3.69)	41.96 (3.24)
ORF-Rus	19.52 (4.68)	9.3 (5.11)	8.33 (2.8)
ORF-Eng	27.9 (8.37)	29.48 (5.3)	28.18 (5.41)

showed the universal lexical effects of length, frequency, and predictability (see Table 2 in Laurinavichyute et al., 2019).

Korneev and colleagues (2017) developed the child version of the RSC for 37 monolingual Russian children (17 girls; $M_{\text{Age}} = 8.6$). The children were second graders in a Moscow public school, and, after one full year of formal literacy instruction, were the youngest group that was able to read in whole words with the average speed of 70 words per minute and comprehend the reading material making them the first age group appropriate for the comparison to adult HSs. The children of this age also represent a particularly interesting group for the comparison with our HS participants with respect to one of the hypotheses in HL acquisition, THE DIVERGENT ATTAINMENT HYPOTHESIS (Benmamoun et al., 2013; Montrul, 2008; Scontras, Fuchs & Polinsky, 2015). It suggests that HL developmental delays, which result from the abrupt switch to the majority language in childhood (typically between the ages of 5 to 9) and inevitable reduced HL exposure, continue into adulthood. Consequently, HSs' language skills including literacy cease to develop beyond the switch point and often resemble those of young monolingual children.

While it was not possible to use the exact same 144 sentences from the RSC, 30 new similar sentences were constructed around the same 30 target nouns from the RSC that were deemed appropriate for 8-year-old children. The child corpus followed 3×2 design, in which authors manipulated the length and frequency of the target nouns. Table 4 (third column) reveals that, as expected, Russian children produced multiple fixations per word with longer mean durations in early and late measures, skipped fewer words, and made more regressive saccades to the currently fixated words. The average mean fixation durations, skipping, and regression rates as well as the effects of word length, word frequency, and their interaction were comparable to what was found for children learning to read in Roman-based alphabets (for review, see Blythe & Joseph, 2011). The effects of length and frequency of the parafoveal words were not investigated.

2.2. The present study: Bilingual Russian Sentence Corpus (BiRSC)

Following the design and analysis in Laurinavichyute et al. (2019), the current study investigates literacy skills of young adult HSs of Russian. First, we establish eye-movement benchmarks in corpus, i.e., reading of the isolated sentences for two proficiency levels (high vs. low). In lieu of the absence of the standardized HL proficiency test which would take into account heritage advantage in some of the linguistic domains (phonology and semantics; oral production and comprehension) and potential disadvantage in others (syntax and morphology; written production and comprehension), we operationally defined proficiency in HL reading as a set of scores in reading speed, quality of reading, and comprehension based on their performance on two reading assessment tests used for monolingual Russian second-graders (Fotekova & Akhutina, 2002). These scores allowed us to classify HSs participants into high-proficient and low-proficient readers (see Method section for details).

Second, we compare HS eye-movement benchmarks to the baseline (i.e., monolingual adults from RSC), and then two comparison groups (i.e., monolingual children from the child RSC and L2 learners tested in the present study). Based on the available research in L2 and HL reading as well as predictions of the weaker links account, we expect eye movements of both groups of HSs to be different from skilled monolingual adults as both groups have reduced HL input, although to a different extent. Nevertheless, we predict that if high-proficient HSs perform

. https://doi.org/10.1017/S136672892000019X

well on two reading assessment tests, they should also perform better (e.g., read faster, skip more words, have shorter fixation durations and regress less) than low-proficient readers or monolingual children and be closer to monolingual adults in their eye-movement characteristics. Accordingly, we expect low-proficient HSs to perform on par with L2 learners, providing support for the weaker links account, and with children, confirming predictions of the divergent attainment hypothesis in HL (Benmamoun et al., 2013; Montrul, 2008).

Finally, in line with the previous findings, we expect to confirm universal lexical effects of length and frequency on the eye movements in reading in HL. However, we hypothesize that low-proficient HSs would show reduced sensitivity to the length and frequency of the parafoveal words focusing more on the currently fixated word (for review of foveal-on parafoveal processing see Drieghe, 2011; for findings in bilingual parafoveal processing, see Whitford & Titone, 2015; 2016).

3. Method

3.1. Participants

Fifty adult English–Russian HSs (23 women, $M_{Age} = 18.4$, range 13–29, $M_{AgeofArrival} = 3.5$) and 27 L2 learners (17 women, $M_{Age} = 21.2$, range 16–43) participated in the study. We recruited participants from two sites, a large urban university and a specialized public high school, both in New York City. The study was approved by the university's Institutional Review Board (IRB) and the New York City Department of Education IRB. Before the start of the study, all participants (over 18 years) or parents/guardians of the participants (under 18 years) signed the informed consent form (minor participants also provided their assent) and filled out the language background questionnaire, administered in English. Bilinguals were matched on the dominant language (English) and age (see Table 2 for other background information).

3.2. Establishing proficiency in Russian HL reading

We wanted to include every HS participant in the study and establish his or her proficiency in Russian using our reading tests as proxy because we lack the standardized assessment. However, recall that many HSs cannot read in Russian at all. To deal with this problem, we included 48 out of total 50 Russian HSs as they minimally matched reading skills of the youngest monolingual group of Russian 8-year-old second graders (Korneev et al., 2017) on the average reading speed (words per minute), quality of reading (syllables vs. words), and text comprehension. To do so, we used two simple reading assessments to approximate their proficiency in Russian HL reading. They were (1) *Russian Word Identification (Word ID-Rus)* task and (2) *Russian Oral Reading Fluency (ORF-Rus)* task.

Russian Word Identification (Word ID-Rus) task

This task was used to screen participants' ability to read the Cyrillic letters because many Russian HSs either cannot read at all or forgot how to read. We adapted the Word Identification task from Fotekova and Akhutina's neuropsychological assessment for elementary school Russian-speaking children (Fotekova & Akhutina, 2002). Participants were asked to read out loud 24 single words in Russian (the complete list of words is presented in Fotekova & Akhutina, 2002, p. 21). The words tested the reader's mastery of the pronunciation of all phonemes as well as of the major phonetic processes in Russian, such as palatalization (SHM**E**l' /ʃm^{j}ˈel'// 'bee'), word final devoicing (FLA**G** /flak/ 'flag'), syllabic stress shift (SAPOG**Í**′$_{PL}$ - SAP**Ó**G$_{SG}$ 'boots-boot'), and vowel reduction (S**O**BÁKA /s**a**báka/ 'dog'). The reading fluency was assessed based on three criteria with a maximum of 15 points for each (45 in total): reading speed (words per minute), method of reading (whole words/syllables/sounds), and number of errors (pronunciation). Oral productions were scored by the native Russian speaker, the first author. The two participants who scored fewer than 15 points in total were classified as not eligible and did not proceed any further in the study.

Russian Oral Reading Fluency (ORF-Rus) task

The remaining 48 participants were asked to read out aloud a short text in Russian *Kak ja lovil rakov* "How I was Catching Crayfish" (202 words; Fotekova & Akhutina, 2002, p. 21). The text utilizes various grammatical constructions (e.g., relative clauses, passives, null object, subject drop, zero copula, impersonal verbs, double negative), tenses (including historic present), different word orders (SVO, VSO, and OVS) embedded into declarative and exclamatory sentences that contained lexical items of different frequencies, illustrated in Example (4).

(4)	Voda	chistaja,	no rakov	ja
	Water$_{NOM\text{-}FEM}$	clean$_{NOM\text{-}FEM}$	but crayfish$_{GEN\text{-}PL}$	I
	ne videl	nigde.		
	not saw$_{PAST}$	nowhere.		

'The water was clean, but I didn't see crayfish anywhere.'

The maximum score for the task was 45 points. We used three criteria for reading fluency: reading speed (words per minute), comprehension score (3 comprehension questions), and number of reading errors (stress, pronunciation, omissions, repetitions). As a result, HSs (and L2) participants were classified into two subgroups, high- and low-proficient readers.

High-proficient HSs ($n = 21$, 13 women, $M_{Age} = 17.5$, range 13–24; $M_{AoA} = 5.1$) scored ≥ 30 pts whereas the remaining 27 HSs were low-proficient (22 women, $M_{Age} = 19$, range 15–29, $M_{AoA} = 2.3$), scoring < 30 pts. All 27 L2 participants were classified as low-proficient readers.

3.3. Design and materials

Having classified HSs according to their proficiency in reading with the help of Word ID-Rus and ORF-Rus tasks, we also administered two parallel tasks in English: namely, *English Word Identification (Word ID-Eng)* and *English Oral Reading Fluency (ORF-Eng)*, to rule out general reading difficulties in their dominant (English) language.

English Word Identification (Word ID-Eng) Subtest

English Word Identification (Word ID-Eng) Subtest of the Woodcock Reading Mastery Tests (WRMT 3rd edition, Woodcock, 2011). Word ID-Eng is a part of the standardized assessment of reading in English. It serves as a test of a decoding skill and requires participants to read out loud English words in the set of 5–6 items of increasing difficulty (e.g., *plausible*, *abdominal* in the initial set and *ennui*, *dossier* in the final set). The task included 17 trials, with a maximum score of 46 points (scoring starts with the baseline of 30 pts). Testing is discontinued after three consecutive errors in

Table 3. Descriptive characteristics of the beginner and advanced versions of BiRSC.

	Beginner	Advanced	
		Version a	Version b
# of sentences	30	72	72
# of words[a]	227	533	541
Sentence length	*M* = 8, range: 6–9	*M* = 9, range: 5–13	
Word length (letters)	*M* = 5.6 *Mdn* = 6, range: 1–13	*M* = 5.7, *Mdn* = 6, range: 1–16	
Word frequency (*ipm*) (# of words)			
Class 1 (1–10)	81	161	181
Class 2 (11–100)	69	150	132
Class 3 (101–1,000)	30	89	83
Class 4 (1,001- 10,000)	24	72	76
Class 5 (10,001-max)	23	61	69

[a]First and last words in each sentence were not analyzed.

pronunciation. All 48 HS participants passed the baseline revealing reading fluency in the dominant language.

English Oral Reading Fluency (ORF-Eng) Subtest

English Oral Reading Fluency (ORF-Eng) Subtest of the Woodcock Reading Mastery Tests (WRMT 3rd edition, Woodcock, 2011). ORF-Eng measures the participant's ability to fluently read connected text. The task is to read out loud sentences that gradually increase in difficulty (e.g., the last sentence was: *Since then, it has been easier with the introduction with the box camera and flexible film, and most recently, the "point-and-shoot" process that requires no specialized knowledge at all.*) Performance was scored for both accuracy and fluency of expression. The experimenter noted errors and omissions, including mispronunciations, word substitutions, hesitations, repetitions, and transpositions. The final score was calculated based on the formula which takes into consideration the total reading time, number of words in the passage, and number of errors. Table 2 provides the oral fluency scores both in Russian and English.

Bilingual Russian Sentence Corpus (BiRSC): Reading experiment

The design and materials for the present study follow the design and materials from RSC (Laurinavichyute et al., 2019) and its child version (Korneev et al., 2017). Low-proficient HSs and L2 learners read the 30 sentences from the child version of RSC whereas high-proficient HSs read one-half of the 144 sentences from RSC (version A the first 72 sentences, version B the second 72 sentences), in order to accommodate time constraints of the study. We refer to A and B versions of RSC as the ADVANCED BiRSC and to its child version as the BEGINNER BiRSC. Table 3 presents the descriptive characteristics of all corpus words and sentences from the beginner and the advanced BiRSC. The sentences and the script used for the analyses reported below are available at the Open Science Framework project page doi:10.17605/OSF.IO/TCRBA.

All words in the BiRSC were annotated for length and frequency. Frequency information was taken from Lyashevskaya and Sharov (2009). Examples (5) and (6) illustrate representative sentences from advanced and beginner versions, respectively. (The morphological markers are omitted for ease of exposition).

To make sure that the participants read for content, all sentences in BiRSC were followed by a multiple-choice comprehension question. High-proficient readers had three alternatives per question (5) and low-proficient readers had two (6).

(5) *Na bolotakh ostavalsya eshchyo lyod,*
'On the marshes remained still ice
no na beregakh reki poyavilas' trava.
but on the banks of the river appeared grass.'
Question: *Chto ostavalos' na bolotakh?*
'What remained on the marshes?'
Multiple-choice: a) *trava* 'grass' b) *lyod* 'ice' c) *tsvety* 'flowers'

(6) *Doroga vela v glukhoj les, petlyaya po sklonam.*
'The road led to the thick forest turning around the slopes.'
Question: *Kuda vela doroga?*
'Where did the road lead?'
Multiple-choice: a) *v gorod* 'to city' b) *v les* 'to forest'

3.4. Procedure

All sentences were presented in Ubuntu Mono Normal black font, size 22 pt, on a light grey background on the BenQ XL2411Z 144 Hz monitor (resolution: 1920 × 1080 pix) controlled by a ThinkStation computer. Presentation was programmed in Experiment Builder (SR Research Ltd.). The eye movements were recorded by the Eyelink 1000+ desktop mount eye-tracker using a chin rest. Participants were seated 55 cm from the camera and 92 cm from the monitor. One letter subtended 0.34° visual angle. Only the right eye was tracked, at 1000 Hz rate.

The experiment began with a 9-point calibration which was repeated after every 15 sentences. Each trial started with the fixation point at the position of the first letter in the sentence presented for 500 ms. If the fixation detection was successful, the experiment automatically proceeded to the presentation of the sentences; otherwise, calibration was repeated. The experiment started with three practice trials. To indicate that they finished reading the sentence, participants fixated the red dot at the lower right-hand corner of the screen; when the fixation was detected by the eye-tracker, the trial proceeded to the comprehension question which the participants answered by clicking on one alternative with a mouse. After 1-s delay, the program proceeded to the next sentence and comprehension question trial. The experimental sentences (72 or 30) appeared on the screen in randomized order.

4. Results

4.1. Descriptive statistics: Reading assessment tasks 1–4

Table 2 (last 4 rows) presents means and standard deviations for reading assessments in Russian and English. The performance in the ORF-Rus was scored based on 1) the pre-defined criteria established by Fotekova and Akhutina (2002), and 2) the Woodcock Reading Mastery Tests formula; the two set of scores were highly correlated, $r = .80$, $p < .001$. We used the latter score in the statistical analyses to ensure the comparability with the score of the ORF-Eng.

The significantly higher scores on both Russian tasks for 21 HSs (Word ID-Rus [t (30)= 5.11, $p = .012$]; ORF-Rus [t (46) = 7.12, $p < .001$]) as compared to 27 remaining HSs and all L2 (Word ID-Rus [t (27) = 4.72, $p < .001$]; ORF-Rus [t (30) = 9.68, $p < .001$] supported the predefined classification of participants as high-proficient and the latter as low-proficient readers. In ORF-Eng, there was no

Table 4. Comparison of basic parameters of eye movements ((i) time duration measures, (ii) probabilities of skipping or fixating the word, (iii) probability of regressions, saccade landing sites and number of fixations per word) in reading in Russian (SD in parentheses).

		Monolingual adults†	High-proficient HSs	Monolingual children	Low-proficient HSs	L2 learners
i (ms)	FF	217 (23)	302 (100)	380 (112)	391 (156)	301 (86)
	SF	228 (26)	305 (73)	358 (78)	345 (164)	312 (71)
	GD	259 (42)	**484 (159)***	676 (273)	944 (307)*	736 (261)
	TT	318 (79)	**702 (221)***	976 (370)	1554 (439)*	1343 (511)
ii (%)	P0	34 (10)	**20 (5)**	20 (13)	10 (5)*	11 (6)
	P1	56 (7)	46 (8)	43 (9)	**37 (8)***	50 (10)
	P2+	9 (6)	34 (10)	35 (11)	**52 (7)***	39 (12)
iii (%)	RO	17 (7)	**23 (9)**	25 (7)	**24 (12)**	36 (15)
	RG	13 (8)	22 (9)*	14 (6)	22 (12)*	25 (13)
Landing (%)		44 (6)	38 (6)	36 (4)	35 (5)	36 (6)
# Fixations		1.01 (.28)	**2.14 (.49)**	2.18 (.65)	4.16 (1.2)*	3.71 (1.4)

†All the differences (RO difference is marginal) are significant between HSs and monolingual adults. * Significant differences between HSs and monolingual children. Significant differences between HSs and L2learners are in bold.

difference among the three groups in the performance, (*ts* < 1). In Word ID-Eng, high- and low-proficient HSs did not differ (t (46) = −1.90, $p = .204$), but L2 readers obtained significantly higher scores than high-proficient HSs (t (46) = −2.88, $p = .018$), suggesting that L2 learners are faster with grapheme-to-phoneme decoding process in English than high-proficient HSs. Low-proficient HSs and L2 readers did not differ on any of the four tasks. (See Table S5 for exact t-values and p-values for all measures, Supplementary Materials).

Performance on the ORF-Rus task was also correlated with the self-reported amount of Russian language exposure per day ($r = .54$, $p = < .001$), age of arrival to the USA ($r = .52$, $p = < .001$), age of reading start in Russian ($r - .51$, $p = <.001$), and self-reported comprehension ability in Russian ($r = .58$, $p < .001$). To make sure that variability in the age range for HSs does not affect the results, we conducted independent t-tests for reading assessments and other background information by splitting HSs into two age groups: adolescents younger than 18 ($n = 28$) and young adults 18 or older ($n = 47$). The only differences (t (73) = 2.08, $p = .041$) was in self-reported age of reading start in Russian ($M_{adolescent} = 9.1$, $SD = 4.7$ vs. $M_{adult} = 12.6$, $SD = 8.2$) and objective measurement for oral reading fluency in English (t (73) = −2.77, $p = .007$) where adolescents scored higher than adults ($M_{adolescent} = 31.1$, $SD = 6.4$ vs. $M_{adult} = 27.1$, $SD = 5.8$).

4.2. Descriptive statistics: eye-movement benchmarks in reading in HL

Table 4 presents means and standard deviations for nine dependent measures from Table 1 for the two groups of high-proficient (second column) and low-proficient HSs (fourth column) and compares them to those of the monolingual Russian adults ($n = 96$, Laurinavichyute et al., 2019), 8-year-old children ($n = 37$, Korneev et al., 2017), and L2 learners ($n = 27$, this study). Additionally, we included the mean number of fixations per word (x) and the saccade landing position (xi). Sentences with incorrect comprehension question responses were excluded from the analysis (low-proficient HSs $M_{accuracy} = 81\%$; high-proficient HSs $M_{accuracy} = 92.7\%$; L2 learners $M_{accuracy} = 85.1\%$; children $M_{accuracy} = 98\%$; monolingual adults $M_{accuracy} = 99\%$). Descriptive statistics were calculated for all words in the corpora, between-group differences were calculated using series of independent t-tests with Bonferroni correction for multiple comparisons at α-level .005 (see Table S6 for t-values and corresponding p-values, Supplementary Materials).

High-proficient HSs

On the one hand, as Table 4 reveals, high-proficient HSs' eye movements were significantly different from eye movements of monolingual adults as assessed with a series of independent t-tests between groups (Table S6). On the other hand, they were strikingly similar to monolingual children in all measures except gaze duration (GD), total reading time (TT) and the probability of regression from the fixated word (RO) (Table 5).

When compared to two low-proficiency groups, low-proficiency HSs and L2, high-proficient HSs were significantly faster in GD and TT measures, skipped more words, re-fixated the words less and, with respect to L2 learners, high-proficient HSs produced lower rates of regressive saccades (RO).

Low-proficient HSs

All low-proficient HSs' eye-movement characteristics were significantly different from those of monolingual adults (RO is marginal). Low-proficient HSs matched children in first fixation duration (FFD) and single fixation duration (SFD) measures (Figure 1A), as well as the probability of regression from the fixated word (RO; Figure 1B) and saccade landing position. However, they fixated more than half of the words at least twice, for longer times (GD and TT), and skipped (P0) fewer words as compared to children. Low-proficient HSs did not differ from L2 learners in any of the measures except fixation probability and regression originating from the fixated word (Table S6).

Next, we compared the differences in frequency effects (FEs) between low-proficient and high-proficient HSs because of the sensitivity to the amount of language exposure in L2 reading (Whitford & Titone, 2012). Words were divided into high frequency (HF) and low frequency (LF) using a median split in log-transformed frequency for high-proficient HSs (HF: $n = 521, \geq 1.49$; LF: $n = 523, < 1.49$) and the low-proficient HS group

Table 5. Conceptual comparison of high-proficient HSs and low-proficient HSs to other groups (to match the statistical analysis in Table 4.) All the differences (RO difference is marginal) are significant between HSs and monolingual adults. Empty cells designate no difference between groups.

		High-proficient HSs			Low-proficient HSs	
		Monolingual children	Low-proficient HSs	L2 learners	Monolingual children	L2 learners
i (ms)	FF	–	–	–	–	–
	SF	–	–	–	–	–
	GD	Shorter	Shorter	Shorter	Longer	–
	TT	Shorter	Shorter	Shorter	Longer	–
ii (%)	P0	–	Higher	Higher	Lower	–
	P1	–	Higher	–	Lower	Lower
	P2+	–	Lower	–	Higher	Higher
iii (%)	RO	Higher	–	Lower	-	Lower
	RG	–	–	–	–	–
Landing (%)		–	–	–	–	–
# Fixations		–	Less	Less	More	–

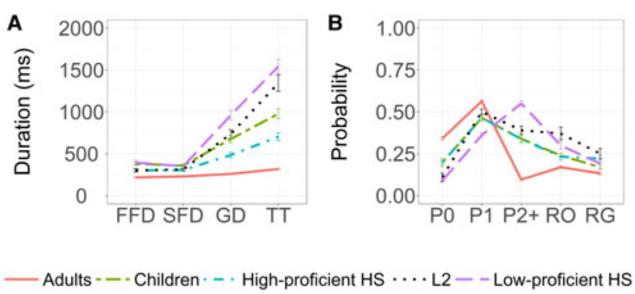

Fig. 1. Means for (A) time durations measures and (B) probabilities of skipping (P0), fixating (P1, P2+) and making regressions (RO, RG) by each group of speakers.

($n = 95, \geq 1.89$; LF: $n = 96, < 1.89$). The median split in advanced BiRSC roughly corresponded to the commonly used thresholds for low (medium)- and high-frequency words (low-frequency range: 1–32 instances per million; high-frequency range: 32–38107 ipm). In the beginner BiRSC, however, the lower frequency range also included words with considerably high ipm count due to the nature of the corpus, in which the sentences were constructed for reading by children (low-frequency range: 1–75 instances per million; high-frequency range: 82–38107 ipm).

Table 6 shows that there are only a few differences in the magnitude of the FEs between the two groups. Low-proficient HSs showed significantly larger FEs in GD, TT, and the probability of making one fixation (P1). The difference between FEs in skipping probability was also significant, but the pattern was reversed: high-proficient HSs skipped frequent words more often than low-proficient HSs. Note that frequency was used as a continuous variable for all subsequent analyses.

4.3. Modeling: Relationships between frequency, length, reading assessments and eye movements in BiRSC

We ran (generalized) linear mixed models using R (R Core Team, 2016) that included previous, current and next words' length and frequency as well as length of incoming saccade to the current word, relative position of the word in the sentence, and saccade landing position on the current word as fixed predictors (see details for the models in Tables S7-S12, Supplementary Materials). Each eye-movement measure was fit to the same set of predictors. Random factors were random intercepts for participants, sentences, and words. No random slopes were added to the final models as such addition resulted in over-parametrization. We also removed random intercepts for sentences in models that resulted in singular fits (the variances across sentences were estimated as zero). Significant effects are adjusted for multiple comparisons using Bonferroni correction at α-level of .005.

To establish the relationship between fixation duration measures in BiRSC and performance for the four reading assessments, we also run a separate set of models that include scores from the reading assessments along with baseline predictors (see details for the models in Tables S13 – S15, Supplementary Materials). Random structure remained the same as in previous set of models. Significant effects are adjusted for multiple comparisons using Bonferroni correction at α-level of .012.

For all analyses, the first and last words of every sentence were excluded and only sentences with correct answers to the comprehension questions were analyzed. Fixations and saccades were extracted from eye-movement data following the algorithm from the Data Viewer package (SR Research Ltd). No cut-off limits were applied to fixations because fixations shorter than 100 ms constituted only 1.9% of all data (the maximum fixation duration was 3779 ms for low-proficient HS that was elicited by low-frequency word of .85 ipm). The predictor of word length was centered and scaled; the frequency was log-transformed (to base 10). Eye-movement duration measures (FFD, SFD, GD, and TT) were log-transformed to ensure normal distribution of models' residuals. To exclude the possibility of multicollinearity of model predictors, we ensured that the variance inflation factor (VIF) was less than 3 for all predictors.

For binary outcome variables (fixation, skipping and regression probability), we used mixed-effects linear logistic regression. Both linear mixed-effects and generalized linear mixed-effects models were fit with the function *(g)lmer* from the R-package lme4 (Bates, Mächler, Bolker & Walker, 2015). The comparison

Table 6. Means and standard deviations for all eye-movement measures across word log frequencies. Frequency effects (FEs; i.e., differences between LF and HF words) and corresponding p-values are presented in italics. First four measures (FFD, SFD, GD, TT) are in ms, the rest are percentages.

	High-proficient HSs			Low-proficient HSs			*t (df)*	*p*
	HF	LF	FE	HF	LF	FE		
FF	266 (73)	316 (112)	*50*	333 (101)	426 (195)	*93*	1.55 (45)	*.128*
SF	284 (64)	332 (90)	*48*	310 (99)	407 (267)	*97*	2.72 (26)	***.011***
GD	349 (91)	593 (221)	*244*	596 (207)	1271 (409)	*675*	14.33 (26)	***<.001***
TT	474 (123)	887 (311)	*413*	964 (311)	2111 (614)	*1147*	8.90 (42.2)	***<.001***
P0	33 (8)	6 (0)	*−27*	13 (5)	6 (6)	*−07*	10.82 (28.7)	***<.001***
P1	49 (7)	43 (13)	*−06*	49 (9)	24 (7)	*−25*	−5.68 (27.2)	***<.001***
P2+	17 (6)	50 (13)	*33*	38 (8)	67 (8)	*29*	−1.17 (45)	*.246*
RO	20 (10)	26 (9)	*06*	20 (12)	28 (13)	*08*	1.40 (45)	*.167*
RG	21 (8)	23 (10)	*02*	19 (10)	24 (14)	*05*	1.68 (45)	*.100*

tables (S7-S14, Supplementary Materials) for (G)LMM outcomes were created with the *sjPlot* package (Lüdecke, 2017).

High-proficient HSs

Confirming canonical effects in reading, the lexical factors of word length and frequency reliably affected all measures of interest (see Tables S7-S8, Supplementary Materials). High-proficient HSs fixated longer words for longer time and frequent words for shorter time, and they regressed to longer words more often and rarely skipped them (see Figure 2).

The effect of the previous (n−1) and upcoming words (n+1)

When the upcoming word was longer, the time spent fixating the current word (GD, TT) decreased. Similarly, higher frequency of the upcoming word decreased total time (TT) spent reading the current word. In regards to the preceding word, longer preceding words led to decreased total reading times (TT) on the current word. When the preceding word had low frequency, high-proficient HSs fixated the current word (TT) longer. In addition, the regression probability decreases when preceding word length and frequency increase (see Tables S7-S8, Supplementary Materials). As effects appeared mostly in total reading time or regression rate measures, we cannot draw any decisive conclusions concerning parafoveal processing, as the word in parafovea could have been fixated before these effects occurred.

Reading assessments

Higher scores in the ORF-Rus were associated with faster reading times for all fixation duration measures (FFD, SFD, GD, and TT). Reading assessments in English did not predict any time duration measures (see Table S13, Supplementary Materials).

Low-proficient HSs

Low-proficient HSs were also sensitive to word length and frequency (see Tables S9-S10, Supplementary Materials) except that the first (FFD) and single fixation duration (SFD) as well as regression rates were not affected by the word length (same finding for SFD for Russian monolingual adults and children). All measures of interest were affected by word frequency (see Figure 3). We did not analyze factors affecting skipping probability as it was very low in general (10%).

The effect of the previous (n−1) and upcoming words (n+1)

Only the regression probability (RG) was affected by the upcoming word frequency wherein more frequent words led to a decreased chance of regression. Longer upcoming words led to reduced first fixation duration (FFD). With respect to the preceding word, longer preceding words led to longer reading times in single fixation duration (SFD). Longer and more frequent preceding words were also associated with lower probability of regressing to previous words (RO) (see Tables S9-S10, Supplementary Materials). Thus, low-proficient were similar to high-proficient HSs: they did process some information parafoveally (e.g., length of the parafoveal word), but most of the effects appeared in the late eye-tracking measures (i.e., regression rates).

Reading assessments

None of the reading assessments in Russian or English predicted any of the time duration measures (see Table S14, Supplementary Materials).

L2 learners

L2 learners were sensitive to word frequency in all duration measures (see Figure 4) except first fixation duration (FFD). Longer words led to increased gaze duration (GD), total reading times (TT), and higher probability to multiple fixations on words (P1, P2+) (see Tables S11-S12, Supplementary Materials).

The effect of the previous (n−1) and upcoming words (n+1)

None of the measures were affected by upcoming or preceding word length or frequency with exception of regression probability (RG), wherein upcoming longer words led to decreased regression rates (see Tables S11-S12, Supplementary Materials).

Reading assessments

None of the reading assessments in Russian or English predicted any of the time duration measures (see Table S15, Supplementary Materials).

5. Discussion

In this eye-tracking study, we presented the *Bilingual Russian Sentence Corpus (BiRSC)* available at the OSF (doi:10.17605/OSF.IO/TCRBA) that characterizes literacy skills by adult and

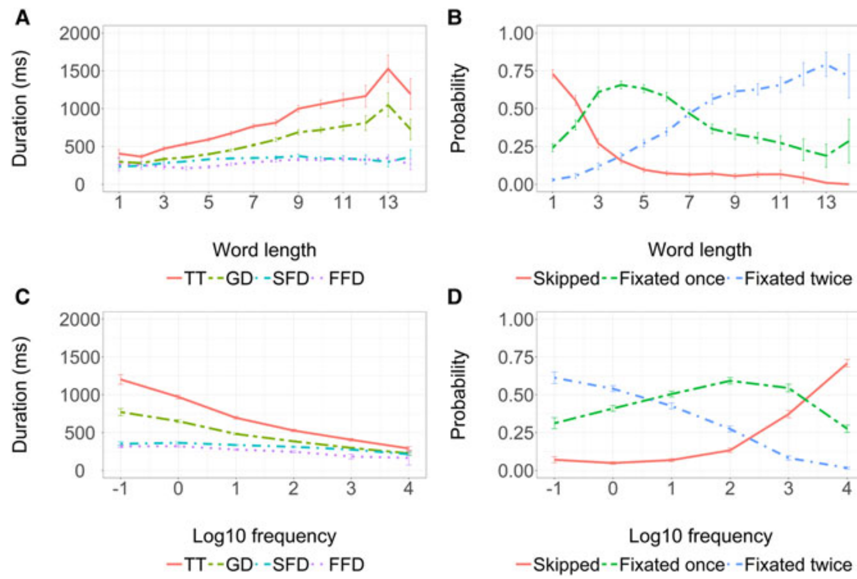

Fig. 2. High-proficient HSs. All corpus words: Means for four time durations measures as a function of word length (A) and logarithmic word frequency (C); probabilities of skipping or fixating the word as a function of word length (B) and frequency (D).

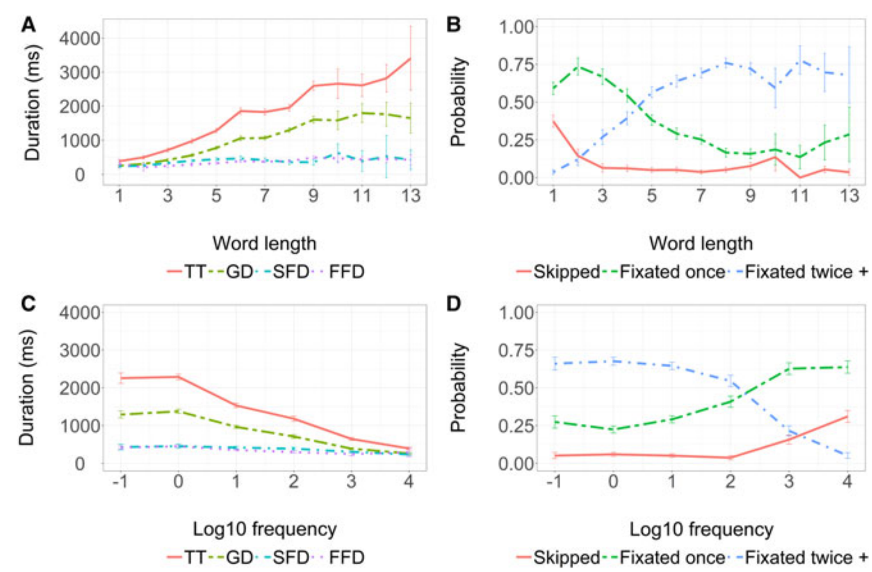

Fig. 3. Low-proficient HSs. All corpus words: Means for four time durations measures as a function of word length (A) and logarithmic word frequency (C); probabilities of skipping or fixating the word as a function of word length (B) and frequency (D).

adolescent HSs of non-Roman-based HL, i.e., Russian. BiRSC contains basic eye-movement characteristics in HL reading as a factor of the HL proficiency level. Then we compared the HSs' eye-movement benchmarks to those of skilled monolingual Russian readers (baseline), children learning to read, and L2 learners. In what follows, we separately discuss the HSs' eye-movement benchmarks and the effects of the lexical characteristics of the words, i.e., length and frequency.

5.1. Proficiency and eye-movement benchmarks in reading in HL

High-proficient HSs were classified as such in our study as those who had high scores on reading assessments tests; these scores were also positively correlated with self-reported amount of Russian language exposure per day, self-reported comprehension ability in Russian, and age of arrival to the USA. Surprisingly, our high-proficient HSs were quantitatively different from the skilled monolingual readers in all of the eye-movement measures. If anything, high-proficient HSs more resembled monolingual children in early duration fixations (except gaze duration), probability of skipping, mean number of fixations, regression rate (probability of making a regression from fixated word), and the saccade landing position. Thus, basic eye-movement benchmarks of high-proficient HSs characterize them as bilingual readers with child-like eye-movement patterns in reading isolates sentences.

Proficiency, however, turned out to have a weak effect on HL reading. It affected only some eye-movement benchmarks: namely, the high-proficient HSs read faster (gaze duration and total reading time), skipped more words, and had fewer fixations than the low-proficiency HSs. However, there were no differences between the two groups in many other measures, i.e., in the earliest (first fixation duration, single fixation duration, saccade landing position) and some of the late eye-movement measures (both saccade rates). We conclude that even proficient HSs experience difficulties in HL during both lexical (early) and post-lexical (late) processing.

One possible limitation of our study is that skilled monolingual adults and high-proficient HSs read one, more difficult, set

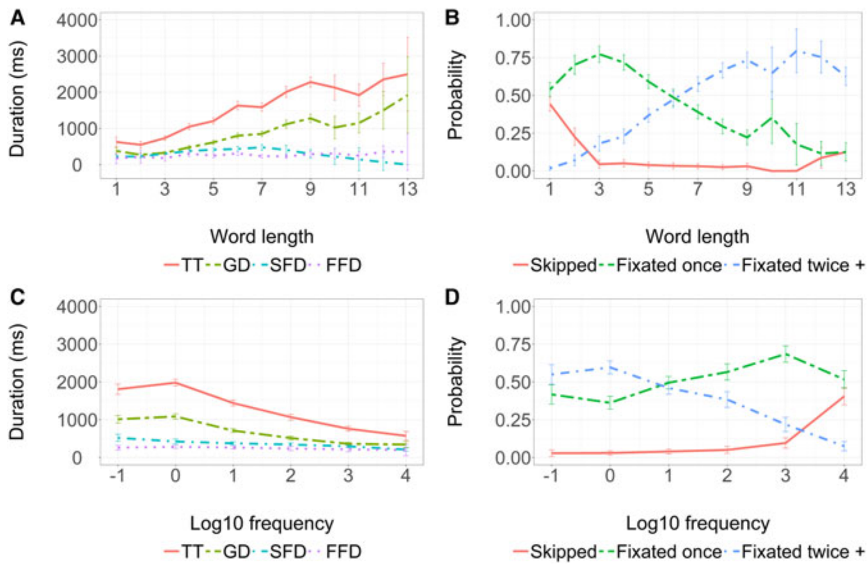

Fig. 4. L2 learners. All corpus words: Means for four time durations measures as a function of word length (A) and logarithmic word frequency (C); probabilities of skipping or fixating the word as a function of word length (B) and frequency (D).

of the sentences whereas children, low-proficient HSs, and L2 learners read another, simpler set of sentences. It was our intention to have the high-proficiency HSs face more complex sentences so that we can compare their eye-movement characteristics to those of adult native speakers and estimate the extent of the gap in reading abilities between these two groups during natural uninterrupted comprehension. The sentences that low-proficiency HSs and L2 learners read were deliberately simplified and were identical to the materials read by the children to allow the direct comparison between children and bilingual readers (recall that children were not able to read the sentences from the RSC). The obtained results of the L2 learners and low-proficient HSs' eye-movement characteristics, therefore, represent a liberal overestimation of their potential performance on the advanced version of BiRSC. The key differences that we found comparing low- to high-proficient group would also hold for more difficult reading materials but with greater dissimilarities between the proficiency levels. As for similarities (in first and single fixation durations, regression rates and saccade landing position), we speculate that while high-proficient HSs would be likely to outperform low-proficient readers, we are now confident that they still do not reach the reading fluency of monolingual adult counterparts. Therefore, we can speak of a continuum of literacy skills in high-proficient HSs that range between "L2 learner" and "monolingual baseline" stages.

So why is it that the reading abilities in even high-proficient HSs in our study differ so much from skilled monolingual readers while resembling monolingual children learning to read more than any other group? We suggest that these differences follow from two theories, one of bilingualism and another of HL. First, as expected, our results support the weaker links account (Gollan et al., 2008) of lexical access delays in bilinguals. Second, the similarities between HSs and monolingual children are consistent with the divergent attainment hypothesis in HL theory (Benmamoun et al., 2013; Montrul, 2008; Polinsky, 2006; Polinsky & Kagan, 2007; Scontras et al., 2015). When HSs switch to the dominant language in childhood, their competence in HL often slows down or even ceases to develop beyond this point. Reduced input in HL and varying exposure to literacy that are difficult to control leave many HSs at the 'child' state of language development even as they reach adulthood.

Turning now to the low-proficient HSs, we found that they were on par with the L2 learners in the majority of the eye-tracking measures but less so with the children. They lagged behind the children (as evident from findings of this study and previous research) in mean gaze duration, total reading times, fixation counts, and skipping probability (Blythe et al., 2011; Blythe & Joseph, 2011). Thus, despite some overlap in early fixation duration measures, the low-proficient HSs group resembles more 'typical' unbalanced L2 learners.

Finally, we would like to point out one important finding that concerns the saccade landing position. Bilingual readers differed from the skilled adult readers in where they first landed the gaze in the word: their saccade landing position was shifted significantly towards the beginning of the word (i.e., 44% for monolingual adults vs. 36% and 38% for L2 learners and HSs, respectively) compared to the expected word-centered OVP (O'Reagan & Jacobs, 1992). In that sense, they were similar to the children (36% into the word). This is a pattern reported for less proficient readers, readers with dyslexia or children who read texts too difficult for their age (e.g., Barnes & Kim, 2016; Hawelka, Gagl & Wimmer, 2010; Kuperman & Van Dyke, 2011). The shift to the beginning of the word signals difficulties in grapheme-phoneme conversion, the process that is automatized in skilled readers. Struggling readers process words in a sequential manner, starting with the beginning and slowly progressing along the word, which leads to multiple refixations. For HSs reading in HL Russian, this grapheme-phoneme conversion could be exacerbated by differences between Cyrillic and Roman scripts, in which only 16 letters out of 33 are shared. We hypothesize that grapheme-phoneme conversion of the weaker HL is inhibited by the dominant language (i.e., English).

5.2. Lexical effects on the eye movements in HL

Our study has confirmed the universal effect of lexical characteristics of the words on eye movements, such as length and frequency (e.g., Inhoff & Rayner, 1986; Staub, White, Drieghe, Hollway & Rayner, 2010) for reading in HL. Regardless of the proficiency level, both groups of HSs showed sensitivity to frequency (in all measures) and length (in all measures in high-proficient HSs, some measures in low-proficient HSs) of the currently fixated words.

One exception is the lack of length effect on first fixation duration and single fixation duration in low-proficient HSs, but it also lacked for single fixation duration in the monolingual adult (Laurinavichyute et al., 2019) and child data (Korneev et al., 2017). Laurinavichyute and colleagues attribute this finding to a specific reading strategy in Russian: namely, that single fixation duration only serves as a quick check for the predictions made for the word before it was fixated, and, therefore, the fixation does not trigger the start of lexical processing and does not depend on the length of the current word. However, we doubt that the same explanation can apply for the low-proficient HSs considering their low skipping probability and high number of fixations per word. The exact same pattern was found in the L2 learners. The reduced lexical access for low-proficient readers might be a more appropriate explanation that led to the chain reaction in the form of low skipping rates and high regression probabilities resulting in multiple and longer fixations (including first fixation) on most words regardless of their length.

The length and frequency of parafoveally presented words had very limited effects on the eye-movement measures in HSs and L2 learners in our study. With a few exceptions, we found that the characteristics of parafoveally presented words had significant impact only on the late eye-movement measures. Specifically, longer and more frequent words in the parafovea decreased the total reading times in high-proficient HSs and reduced regression rates on the currently fixated word in all bilingual groups. However, these findings do not allow us to make strong conclusions concerning parafoveal processing in HSs as the effects occurred mostly beyond the first pass reading (i.e., parafoveal words can receive fixations in second, third etc. passes). This pattern of results suggests that while HSs and L2 learners show some sensitivity to the lexical characteristics of parafoveal words, parafoveal processing has little impact on the initial stages of lexical access of the currently fixated word.

Finally, the frequency effects for HL reading were partially confirmed in our data. According to the weaker links account (Gollan et al., 2008), the connection between word forms and their mental representation is weakened due to the reduced exposure to the bilinguals' languages. Accordingly, bilinguals show larger frequency effects in their weaker language relative to monolinguals or relative to reading in their dominant language (Gollan, Slattery, Goldenberg, Van Assche, Duyck & Rayner, 2011; Whitford & Titone, 2012; cf. Cop et al., 2015; Duyck, Vanderelst, Desmet & Hartsuiker, 2008). The weaker links account also suggests that less proficient bilinguals should show larger frequency effects in the non-dominant language compared to more proficient bilinguals. Our findings are only partially consistent with this prediction; although the low-proficient HSs showed numerically larger frequency effects, significant differences were found only for gaze duration, total reading time, and the probability of fixating the word only once. It is plausible, therefore, that at the earliest stages of lexical access (reflected in single fixation and first fixation durations, skipping probability), greater proficiency in HL does not lead to more efficiency although it plays some role during later stages of HL processing.

In conclusion, this is the first study that has investigated and described basic eye-movement benchmarks in reading in heritage language and compared them to those of monolingual skilled readers, 8-year-old children learning to read, and L2 learners. Our findings suggest that although the proficiency level in HL reading had some effect on the early and late eye-movement measures in reading isolated sentences, its effect was limited. Both high- and low-proficient HSs were more similar to monolingual children learning to read than to skilled readers. Low-proficient HSs were at a particular disadvantage and resembled unbalanced L2 learners more than any other group, suggesting that early exposure to spoken HL does not seem to facilitate literacy and reading fluency in heritage language. We hope the findings reported here will serve as the first step for future research on reading in the heritage language field. Taken into account high variability in spoken HL skills of heritage speakers, an investigation of HSs' individual differences in reading constitutes the next logical step in systematic study of heritage languages.

Supplementary Material. For supplementary material accompanying this paper, visit http://dx.doi.org/10.1017/S136672892000019X

List of supplementary materials
Table S5
Table S6
Tables S7
Table S8
Table S9
Table S10
Table S11
Table S12
Table S13
Table S14
Table S15

Acknowledgements. The contribution of Anna K. Laurinavichyute and Irina A. Sekerina was funded within the framework of a subsidy by the Russian Academic Excellence Project '5–100'. Irina A. Sekerina was also supported by a PSC-CUNY award TRADB-48–172. We are grateful to Aleksei Korneev for providing the data for monolingual children and Svetlana Alekseeva for the help with corpus annotation process. All errors are our responsibility.

References

Barnes AE and Kim YS (2016) Low-skilled adult readers look like typically developing child readers: a comparison of reading skills and eye movement behavior. *Reading and Writing* **29**, 1889–1914.
Bates D, Mächler M, Bolker B and Walker S (2015) Fitting linear mixed-effects models using lme4. *Journal of Statistical Software* **67**, 1–48.
Benmamoun E, Montrul S and Polinsky M (2013) Heritage languages and their speakers: Opportunities and challenges for linguistics. *Theoretical Linguistics* **39**, 129–181.
Berzak Y, Katz B and Levy R (2018) Assessing language proficiency from eye movements in reading. In Walker MA, Ji H & Stent A (eds), *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human language technologies*. New Orleans, LA: Association for Computational Linguistics, pp. 1986–1996.
Blythe HI and Joseph H. S. S. L. (2011) Children's eye movements during reading. In Liversedge SP, Gilchrist ID & Everling S (eds), *The Oxford handbook of eye movements*. New York, NY: Oxford University Press, pp. 643–662.
Blythe HI, Häikiö T, Bertam R, Liversedge SP and Hyönä J (2011) Reading disappearing text: Why do children refixate words?. *Vision Research* **51**, 84–92.
Boston MF, Hale J, Kliegl R, Patil U and Vasishth S (2008) Parsing costs as predictors of reading difficulty: An evaluation using the Potsdam Sentence Corpus. *Journal of Eye Movement Research* **2**, 1–12.
Carreira M and Kagan O (2011) The results of the National Heritage Language Survey: Implications for teaching, curriculum design, and professional development. *Foreign Language Annals* **44**, 40–64.
Clifton C, Jr., Staub A and Rayner K (2007) Eye movements in reading words and sentences. In van Gompel RPG, Fischer MH, Murray WS & Hill RL

(eds), *Eye movements: A window on mind and brain*. Amsterdam: Elsevier, pp. 341–369.

Cop U, Drieghe D and Duyck W (2015) Eye movement patterns in natural reading: A comparison of monolingual and bilingual reading of a novel. *PLoS ONE* **10**, e0134008. doi:10.1371/journal.pone.0134008

Cop U, Dirix N, Drieghe D and Duyck W (2017) Presenting GECO: An eye-tracking corpus of monolingual and bilingual sentence reading. *Behavior Research Methods* **49**, 602–615.

Drieghe D (2011) Parafoveal-on-foveal effects in eye movements during reading. In Liversedge SP, Gilchrist ID and Everling S (eds), *Oxford handbook on eye movements*. Oxford: University Press, pp. 839–855.

Duyck W, Vanderelst D, Desmet T and Hartsuiker RJ (2008) The frequency effect in second–language visual word recognition. *Psychonomic Bulletin & Review* **15**, 850–855.

Engbert R, Nuthmann A, Richter EM and Kliegl R (2005) SWIFT: A dynamical model of saccade generation during reading. *Psychological Review* **112**, 777–813. doi:https://doi.org/10.1037/0033-295X.112.4.777

Fotekova TA and Akhutina TV (2002) *Diagnostika rečevyx narušenij škol'nikov s ispol'zovaniem nejropsixologičeskix metodov* [In Russian. Diagnosis of speech disorders by neuropsychological method]. Moscow: Arkti.

Gollan TH, Montoya RI, Cera C and Sandoval TC (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 TH, Slattery TJ, Goldenberg D, Van Assche E, Duyck W and Rayner K (2011) Frequency drives lexical access in reading but not in speaking: The frequency–lag hypothesis. *Journal of Experimental Psychology: General* **140**, 186–209.

Gor K, Cook SV, Pandza NB and Chrabaszcz A (2018, February). *Hearing is not seeing: Is there an auditory advantage and a visual disadvantage in Heritage Speakers compared to late L2 learners?*. Paper presented at the Third International Conference on Heritage/Community Languages, Los Angeles, CA.

Häikiö T, Bertram R, Hyönä J and Niemi P (2009) Development of the letter identity span in reading: Evidence from the eye movement moving window paradigm. *Journal of Experimental Child Psychology* **102**, 167–181.

Hawelka S, Gagl B and Wimmer H (2010) A dual–route perspective on eye movements of dyslexic readers. *Cognition* **115**, 367–379.

Huey EB (1908) *Basic studies on reading*. New York, NY: Basic Books.

Hyönä J and Bertram R (2011) Optimal viewing position effects in reading Finnish. *Vision Research* **51**, 1279–1287.

Inhoff AW and Rayner K (1986) Parafoveal word processing during eye fixations in reading: Effects of word frequency. *Perception & Psychophysics* **40**, 431–439.

Joseph HS, Liversedge SP, Blythe HI, White SJ and Rayner K (2009) Word length and landing position effects during reading in children and adults. *Vision Research* **49**, 2078–2086.

Ke C (1998) Effects of language background on the learning of Chinese characters among foreign language students. *Foreign Language Annals* **31**, 91–102.

Keating GD, VanPatten B and Jegerski J (2011) Who was walking on the beach?: Anaphora resolution in Spanish heritage speakers and adult second language learners. *Studies in Second Language Acquisition* **33**, 193–221.

Kliegl R, Grabner E, Rolfs M and Engbert R (2004) Length, frequency, and predictability effects of words on eye movements in reading. *European Journal of Cognitive Psychology* **16**, 262–284.

Koda K, Zhang Y and Yang CL (2008) Literacy development in Chinese as a heritage language. In He AW & Xiao Y (eds), *Chinese as a heritage language: Fostering rooted world citizenry*. Honolulu: Foreign Language Annals, National Foreign Language Resources Center, pp. 137–149.

Korneev A, Akhutina T and Matveeva E (2017) Silent reading in Russian junior schoolchildren: an eye tracking study. *Psychology, Journal of Higher School of Economics* **14**, 219–235.

Kuperman V and Van Dyke JA (2011) Effects of individual differences in verbal skills on eye–movement patterns during sentence reading. *Journal of Memory and Language* **65**, 42–73.

Laurinavichyute AK, Sekerina IA, Alexeeva S, Bagdasaryan K and Kliegl R (2019) Russian Sentence Corpus: Benchmark measures of eye movements in reading in Russian. *Behavior research methods* **51**, 1161–1178.

Li X, Liu P and Rayner K (2011) Eye movement guidance in Chinese reading: Is there a preferred viewing location?. *Vision Research* **51**, 1146–1156.

Liversedge SP, Drieghe D, Li X, Yan G, Bai X and Hyönä J (2016) Universality in eye movements and reading: A trilingual investigation. *Cognition* **147**, 1–20.

Lüdecke D (2017) sjPlot: Data visualization for statistics in social science (R package version 2.3.3). Retrieved from https://CRAN.Rproject.org/package=sjPlot

Lyashevskaya ON and Sharov SA (2009) *Chastotnyj Slovar' Sovremennogo Russkogo Jazyka* (*na Materialakh Natsional'nogo Korpusa Russkogo Jazyka*) [*Frequency Dictionary of Modern Russian (based on the materials of the Russian National Corpus)*]. Moscow, Russia: Azbukovnik.

Mancheva L, Reichle ED, Lemaire B, Valdois S, Ecalle J and Guérin-Dugué A (2015) An analysis of reading skill development using E-Z Reader. *Journal of Cognitive Psychology* **27**, 657–676.

Montrul SA (2008) *Incomplete acquisition in bilingualism: Re–examining the age factor*. Amsterdam: John Benjamins Publishing.

Nuthmann A, Engbert R and Kliegl R (2005) Mislocated fixations during reading and the inverted optimal viewing position effect. *Vision research* **45**, 2201–2217.

Paterson KB, Almabruk AA, McGowan VA, White SJ and Jordan TR (2015) Effects of word length on eye movement control: The evidence from Arabic. *Psychonomic bulletin & review* **22**, 1443–1450.

Rakhlin N, Kornilov SA and Grigorenko EL (2017) Ch. 16 Learning to read Russian. In Verhoeven L & Perfetti C (eds), *Learning to read across languages and writing systems*. Cambridge, UK: Cambridge University Press, pp. 371–392.

O'Regan JK and Jacobs AM (1992) Optimal viewing position effect in word recognition: A challenge to current theory. *Journal of Experimental Psychology: Human Perception and Performance* **18**, 185–197.

Polinsky M (2006) Incomplete acquisition: American Russian. *Journal of Slavic Linguistics* **14**, 191–262.

Polinsky M (2018) *Heritage languages and their speakers*. Cambridge, UK: Cambridge University Press.

Polinsky M and Kagan O (2007) Heritage languages: In the 'wild' and in the classroom. *Language and Linguistics Compass* **1**, 368–395.

Potowski K, Jegerski J and Morgan-Short K (2009) The effects of instruction on linguistic development in Spanish heritage language speakers. *Language Learning* **59**, 537–579.

Radach R and Kennedy A (2004) Theoretical perspectives on eye movements in reading: Past controversies, current issues, and an agenda for future research. *European Journal of Cognitive Psychology* **16**, 3–26.

R Core Team. (2016) *R: A language and environment for statistical computing [Computer software manual]*. Vienna, Austria. Retrieved from https://www.R-project.org/

Rayner K (1998) Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin* **124**, 372–422.

Rayner K (2009) Eye movements and attention in reading, scene perception, and visual search. *The Quarterly Journal of Experimental Psychology* **62**, 1457–1506.

Reichle ED, Pollatsek A, Fisher DL and Rayner K (1998) Toward a model of eye movement control in reading. *Psychological Review* **105**, 125–157. doi:https://doi.org/10.1037/0033-295X.105.1.125

Roberts L and Siyanova–Chanturia A (2013) Using eye–tracking to investigate topics in L2 acquisition and L2 processing. *Studies in Second Language Acquisition* **35**, 213–235.

Scontras G, Fuchs Z and Polinsky M (2015) Heritage language and linguistic theory. *Frontiers in Psychology* **6**. doi:10.3389/fpsyg.2015.01545

Share D (2008) On the Anglocentricities of current reading research and practice: The perils of overreliance on an "outlier" orthography. *Psychological Bulletin* **134**, 584–615.

Staub A, White SJ, Drieghe D, Hollway EC and Rayner K (2010) Distributional effects of word frequency on eye fixation durations. *Journal of Experimental Psychology: Human Perception and Performance* **36**, 1280–1293.

Staub A and Rayner K (2007) Eye movements and on–line comprehension processes. In Gaskell MG (ed), *The Oxford handbook of psycholinguistics*. Oxford: Oxford University Press, pp. 325–342.

Tse L (2001) Heritage language literacy: A study of US biliterates. *Language Culture and Curriculum* **14**, 256–268.

Valdés G (2000) The teaching of heritage languages: An introduction for Slavic–teaching professionals. In Kagan O & Rifkin B (eds), *The learning and teaching of Slavic languages and cultures*. Bloomington, IN: Slavica Publishers, pp. 375–403.

Vasishth S, von der Malsburg T and Engelmann P (2013) What eye movements can tell us about sentence comprehension? *Wiley Interdisciplinary Review* **4**, 125–134.

Vitu F, McConkie GW, Kerr P and O'Regan JK (2001) Fixation location effects on fixation durations during reading: An inverted optimal viewing position effect. *Vision research* **41**, 3513-3533.

Whitford V and Titone D (2012) Second–language experience modulates first–and second–language word frequency effects: Evidence from eye-movement measures of natural paragraph reading. *Psychonomic Bulletin & Review* **19**, 73–80.

Whitford V and Titone D (2015) Second-language experience modulates eye movements during first-and second-language sentence reading: Evidence from a gaze-contingent moving window paradigm. *Journal of Experimental Psychology: Learning, Memory, and Cognition* **41**, 1118 –1129.

Whitford V and Titone D (2016) Eye movements and the perceptual span during first–and second–language sentence reading in bilingual older adults. *Psychology and Aging* **31**, 58–70.

Whitford V and Titone D (2017) The effects of word frequency and word predictability during first–and second–language paragraph reading in bilingual older and younger adults. *Psychology and Aging* **32**, 158–177.

Whitford V and Joanisse MF (2018) Do eye movements reveal differences between monolingual and bilingual children's first-language and second-language reading? A focus on word frequency effects. *Journal of experimental child psychology* **173**, 318–337.

Woodcock RW (2011) *Wookcock reading mastery tests third edition manual*. Oxford: Pearson.

Xiao Y (2006) Heritage learners in the Chinese language classroom: Home background. *Heritage Language Journal* **4**, 47–56.

Yarbus AL (1967) *Eye movements and vision*. New York, NY: Plenum Press.

Zhang H and Koda K (2018) Vocabulary knowledge and morphological awareness in Chinese as a heritage language (CHL) reading comprehension ability. *Reading and Writing* **31**, 53–74.