DiSlaw
Didaktik slawischer Sprachen

Maria Dorbert, Humboldt-Universität zu Berlin, Germany

Lexical and spelling accuracy of Russian adult heritage and foreign language learners in written elicited narratives

A Preliminary Study

This paper presents a preliminary study on correlations between the frequency of lexical and orthographic errors in written narratives produced by learners of Russian as a foreign language (FLLs; n = 11) and heritage speakers of Russian (HLLs; n = 14), in two elicitation modes: telling and retelling. The study also explores the correlation between lexical accuracy and overall language proficiency in these two groups. Narratives were elicited using the Multilingual Assessment Instrument for Narratives (MAIN), and a validated multiple-choice language proficiency test served as the anchor test. The results of a preliminary study based on a small sample of participants showed that Russian FLLs produced significantly more lexical errors and fewer orthographic errors in both retelling and telling modes compared to HLLs. Moreover, a significant negative correlation was found between the frequency of lexical errors in telling and the results of the language proficiency test in HLLs. These findings show that the relationship between different components of language competence may differ significantly between HLLs and FLLs and be shaped by different patterns. The study further emphasizes the need for individual instructional strategies for different types of language learners.

Keywords: SLA, heritage speakers, language proficiency, lexical errors, spelling errors

В статье представлены результаты предварительного исследования, нацеленного на изучение взаимосвязи между частотой лексических и орфографических ошибок в письменных нарративах изучающих русский язык как иностранный (РКИ; n = 11) и эритажных носителей русского (ЭН; n = 14), в формате рассказа и пересказа. Исследование также оценивает корреляцию между параметром лексической грамотности и общим уровнем языковой компетенции (ЯК) в каждой из групп участников. Нарративы собраны с использованием многоязычного инструмента оценки нарративов (MAIN), а для оценки уровня ЯК использовался закрытый тест на знание русского. Результаты предварительного исследования на небольшой выборке респондентов показали, что изучающие РКИ допускают значительно больше лексических ошибок и меньше орфографических в формате как пересказа, так и в рассказа, чем ЭН. При сопоставлении частоты лексических ошибок в формате рассказа и результатов теста на ЯК в группе ЭН была обнаружена значимая отрицательная корреляция. Таким образом, исследование обращает внимание на то, что взаимосвязь между разными компонентами ЯК у изучающих РКИ и ЭН может различаться, а также ещё раз подчёркивает необходимость применения разных стратегий преподавания языка вышеупомянутым группам студентов.

Ключевые слова: русский как иностранный, эритажные носители, уровень владения языком, лексические ошибки, орфографические ошибки

DiSlaw
Didaktik slawischer Sprachen

1. Introduction

In the last decades, narratives have become a prominent and widely used approach for evaluating the language abilities of bi-/multilingual speakers. Narratives have been recognized as an effective tool for assessing linguistic skills, as they account for speaker diversity and can be applied across all proficiency levels, including illiterate speakers (Flecken, 2011; Gutiérrez-Clellen et al., 2000). Moreover, narratives are relatively simple to elicit, requiring minimal resources while still offering a reliable set of linguistic data (Pavlenko, 2008).

In educational settings, narratives that are elicited in a target language (TL) can serve as valuable material for assessing different aspects of the language proficiency (LP) on the microstructural level, establishing errors and non-canonical patterns that a language learner (LL) produces, and identifying the individual needs of diverse LLs. This is particularly relevant when teaching language to a mixed group of learners that presents the contemporary reality of language teaching at the university level across various countries with high immigration rates. To date, not many universities offer separate tracks for heritage language learners (HLLs) and foreign language learners (FLLs) due to administrative challenges (for the situation in the USA, see Beaudrie, 2011; in Germany, Ermakova, 2023). As for those institutions in which separate tracks are available, language instructors are often unprepared to address their unique linguistic needs due to a lack of pedagogical training in heritage language (HL) education.

HLLs and FLLs show differences in their language skills as well as in various sociolinguistic aspects, and their linguistic needs vary greatly (Montrul, 2012; 2023; Polinsky, 2015, a.o). HLLs, in turn, are themselves heterogeneous. These speakers show considerable variability in their language skills depending on factors such as the quantity of exposure, family language use (Vorobyova & Bel, 2021) or age of acquisition of the TL (Jia & Aaronson, 2003; Montrul, 2008, 2016; Vorobyova & Bel, 2021) as well as genetic, physiological, cognitive, developmental, and environmental factors (Sekerina, 2013). Some HLLs show native-like LP, whereas some of them display fluency in oral speech, but have poor literacy skills (cf. Polinsky, 2015). Some HLLs can understand the language, but their production skills are limited.

To date, research in the area of HLs remains unevenly developed. On the one hand, there is still a lack of studies that compare language abilities of HLLs and FLLs. Most research focuses on a "heritage vs. baseline speaker" opposition. However, in educational settings, comparing the language of HLLs with that of FLLs often holds greater practical relevance. Given that HLLs, in general, have implicit knowledge in different language domains, some topics that are significant for FLLs do not need to be addressed in detail for HLLs. Thus, research based on a detailed comparison of language abilities of HLLs and FLLs could help develop unified approaches for differentiated and more effective teaching in mixed groups.

On the other hand, much is already known about the general profile of heritage speakers (Montrul, 2012, 2018; Polinsky, 2015, a.o.); phonology (Kupisch et al., 2014), morphology (Montrul, 2004; Polinsky, 2006a), syntax (Benmamoun et al., 2013; Laleko, 2010; Montrul, 2008; O'Grady et al., 2001; Polinsky, 2008; 2010), semantics (Polinsky, 1995; 2006b), and pragmatics (Dubinina et. al, 2010; Sorace & Serratrice, 2009). However, research on the lexicon, especially regarding the Russian language, remains significantly limited.

In Second Language Acquisition (SLA), the accurate use of vocabulary is an integral part of overall language competence. Lexical accuracy, as one of the measures of lexical competence in LLs, is included in widely used research models, such as the complexity, accuracy, and fluency (CAF) model (Skehan, 1988) or the model of lexical richness (Read, 2000). Despite extensive research on this parameter in the context of mostly English language, the relationship between lexical accuracy and overall language competence has been relatively underexplored in the study of Russian FLLs and even more so in the context of Russian as a HL, although it has been widely used by both educators and researchers as a component of language competence assessment in HL (Vorobyova & Bel, 2021) and SLA (Krasnoschekova & Kashleva, 2019) contexts. Moreover, the question of whether the writing task mode (telling or retelling) affects the frequency of lexical errors remains open.

Taking these gaps into account, the present study, on the one hand, aims to contribute to HL didactics and theory by examining how HLLs and FLLs differ in lexical error frequency. On the other hand, it seeks to add knowledge to narrative theory by investigating whether the frequency of lexical errors changes for the same LLs across different modes (telling and retelling). Finally, the study expands the understanding of Russian language proficiency assessment by exploring the relationship between language proficiency scores and lexical error frequency in written narratives.

1.1 Lexical accuracy and its borders

Housen & Kuiken (2009, 461) defined accuracy in general as "the ability to produce error-free speech". However, lexical accuracy remains one of the most challenging aspects to evaluate, as "the line between lexical and grammar errors is blurred" (Hemchua & Schmitt, 2006, 70). Moreover, the taxonomies developed by different researchers for various purposes, often not described in detail and varying significantly across studies, make the assessment of lexical accuracy less transparent and the results difficult to compare directly. The broad range of approaches is well reflected in the following definition, which is based on general trends and different angles traced in research:

> a 'lexical error' is a deviation in form and/or meaning of a target-language lexical word. Form deviations include orthographic or phonological deviations within the limits of single words, and also ignorance of syntactical restrictions which result in false collocations, for example. Meaning deviations appear when lexical items are used in contexts where they are attributed another meaning or where they violate semantic restrictions'; in Berkoff's words (1982, 10), when there is the 'incorrect choice of lexical items', or also when some semantic feature is not considered (Llach, 2011, 75).

To minimize a potential lack of transparency in distinguishing between error types, the present study adopts the framework of Error Analysis (EA; Corder, 1974; Ellis, 1994), following the hierarchical structure of error classification (Hoffmann, 2022), which means that errors should be initially categorized by type (e.g. spelling, lexical, grammatical). This is particularly relevant given that the boundaries between error types are often not so strict: not only between lexical and grammatical errors, but also between lexical and orthographic/phonological ones. Some researchers consider errors in spelling as lexical ones, while others assess them separately (Leńko-

Szymańska, 2019). Given that HLLs are often noted to face significant difficulties with spelling and orthography (Brüggemann, 2018; Kagan & Dillon, 2001; a.o.), the current study distinguishes between lexical and spelling errors.

1.2 Multilingual Assessment Instrument for Narratives: Microstructural assessment in the lexical domain

Narratives can be elicited and evaluated by different techniques, one of them being storytelling based on pictures (Karlsen et al., 2016; Schneider et al., 2006; a.o.). One of the most widespread instruments is the wordless picture book "Frog, Where Are You?" (Mayer, 1969). Another instrument that allows fictional narrative elicitation and assessment and is world-used[1] is the Multilingual Assessment Instrument for Narratives (hereafter MAIN; Gagarina et al., 2012, 2019a). MAIN is a tool originally designed for bilingual preschoolers and primary-school children. Later, it has also been used to assess older children, adolescents, and adults (e.g. Gagarina et al., 2019b). MAIN can be used to evaluate the comprehension and production of narratives in three elicitation modes (telling, retelling, model story) and to assess macro- and microstructure.[2] Macrostructure has been found to be language-independent (e.g. Pearson, 2002) and has a validated scoring scheme. Microstructure is considered language-dependent, requiring the development of tailored assessment scales for each language and specific purposes (e.g. educational settings, diagnosing DLD in children, etc.).

The first and to date single attempt to create an overarching microstructure assessment scale and to examine the relationship between narrative competence and language proficiency of adult Russian second language (L2) learners (mostly Chinese, Vietnamese, and Mongolian speakers) using the MAIN was made by Krasnoshchekova & Kashleva (2019). The authors outlined four sections for the microstructural assessment via error-based analysis: morphosyntax, reference, complexity, and vocabulary. Their scoring sheet for lexical assessment considered four positions: (a) "usage mistakes (wrong meaning)"; (b) "usage mistakes (non-existing words)"; (c) "words from other languages (code-switching)"; and (d) total number in tokens of internal state terms (IST)[3] as a separate measure. Regarding the IST, the perspective on whether it should belong to the microstructural-level of narrative or the macrostructural one is quite ambiguous. According to the MAIN authors, IST is interpreted as a marker of the participant's "understanding and awareness of intentionality and goal-directed behaviour of protagonists" (Gagarina et al., 2012), thus it is closely related to the macrostructure. Moreover, the number of IST has been found to be language-independent, unlike the microstructure of narratives.

In terms of lexicon, Krasnoshchekova & Kashleva (2019) found that although the L2 narratives, regardless of the language level (A2, B1 or B2), are significantly different from the first language (L1) narratives by vocabulary in total, wrong meaning of words, and non-existent words, in L2 narratives these parameters do not change from level to level. The authors suppose that the

---

[1] To date, MAIN exists in 100 language versions from 66 countries (ZAS).

[2] Macrostructure refers to the overarching hierarchical organization of narrative texts, such as episodic structure and story grammar components (Heilmann et al., 2010). Microstructure involves the internal linguistic elements that contribute to cohesive discourse.

[3] IST describe the internal state of the protagonist as a response to the initiating event.

reason for such results may be explained by "the method itself which is quantitative" (*ibid*, 631). However, lexical development at different levels of proficiency still requires further detailed investigation.

1.3 The current study

Expanding the existing knowledge on the relationship between lexical and spelling errors and their connection to general language proficiency, the current study focuses on the following research questions:

1. How does the frequency of lexical and spelling errors vary within task mode (telling vs. retelling) and across the LL groups (FLLs vs. HLLs)?
2. How does the error type (lexical vs. spelling) correlate with participant's language proficiency test score in (a) HLLs, and (b) FLLs, in each narrative mode?

2. Method

2.1 Ethics statement

Ethical approval for this study was obtained from the Ethics Committee of the Faculty of Language and Literature, Humboldt University of Berlin.[4] Before participation, all participants signed informed consent.

2.2 Participants

Participants were invited to take part in the experiment based on the following criteria: healthy young adults aged 18–30 who (a) were either born in or relocated to a German-speaking country before the age of 8 and currently resided there; (b) were university students enrolled in Russian language courses and/or pursuing B.A. or M.A. programs in Teaching of Russian, Russian-German translation or Slavic studies; and (c) self-assessed their Russian language proficiency at the B2 or C1 level. As a result, the participants are highly heterogeneous in terms of language skills and cultural backgrounds, so the study takes into account a wider range of possible variations in language production than it would if the data were collected from a group of learners with the same language instructor within a single institution.

---

[4] SLF-Ethikkommission, Humboldt-Universität zu Berlin, Ethikantrag 2023-4.

2.3 Materials

2.3.1 Language proficiency assessment

To date, there is only one peer-reviewed publication describing a validated test[5] for the Russian LP assessment in both FLLs and HLLs (Luchkina et. al, 2021). For this study, it was adapted in the multiple-choice format and programmed in Javascript code based on the jsPsych library on the online platform *cognition.run.* All orthographic and punctuation inaccuracies and errors found in questions nr. 2, 37 and 55 were carefully corrected.

2.3.2 Narrative tests

Narrative production was elicited using the "Baby Birds" (retelling mode, Stimulus 1) and "Cat" (telling mode, Stimulus 2) stories from the MAIN (Gagarina et al., 2012, 2019a). The research was designed during the COVID-19 pandemic; therefore, the Hong Kong Polytechnic University and ZAS Version of MAIN PPT (ZAS), which was one of the primary adaptations of the original in-person MAIN testing format for remote administration, was used in this study. Being originally designed for online testing with children in an oral format, the template created by the Hong Kong Polytechnic University and ZAS included a picture of a girl as a virtual listener to whom children were asked to tell their stories. This was designed to reduce the shared knowledge be-tween the speaker and the listener. In the present study, in which written narratives from adults were collected, this image was deleted. Due to the written format of narrative production, par-ticipants were instead instructed to produce highly detailed narratives, as they were told that the reader would be someone who had not seen the pictures. Additionally, this instructor en-couraged participants to pay more attention to details, potentially expanding the range of vo-cabulary used in their narratives.
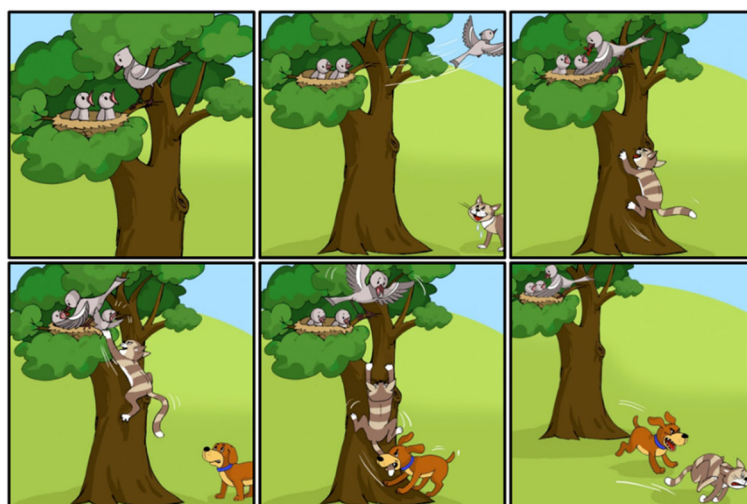
Stimulus 1: Baby Birds (MAIN); Source: ZAS

---

[5] The test was developed and validated in both cloze and multiple-choice formats.

Stimulus 2: Cat (MAIN); Source: ZAS

2.3.3 Procedure

The first part of the experiment (approximately 15 min.) consisted of (1) online questions about language background based on the abridged version of the Bilingual LEAP Questionnaire (LEAP-Q, Marian et al., 2007) designed to differentiate between HLLs and FLLs, and (2) a language proficiency test in a multiple-choice format (Luchkina et al., 2021), both programmed in Javascript code based on the jsPsych library on the platform *cognition.run*. Those participants who demonstrated Level 3 or 4 results according to an adapted version of the language proficiency test (Luchkina et al., 2021) and expressed a willingness to take part further, were invited to participate in the second stage of the experiment. The second stage of the experiment was scheduled within a few days (no more than a week) after the first part, at a time that was convenient for each participant individually. Before the start of the second part, the participant ensured that s/he had (1) an Internet connection and a computer or an iPad with a webcam and a Zoom account on it, (2) a quiet space available for the duration of the assessment, (3) paper and pen and the ability to take/send a photo of hand-written narratives.

The second stage of the experiment was conducted online during a Zoom meeting following the MAIN Protocol. After a brief oral conversation in Russian, the tasks were completed in the following order: (1) Writing Story 1 ("Baby Birds") based on Stimulus 1 after listening to the instructor read the story (retelling mode); (2) Orally answering comprehension questions about Story 1; (3) Writing Story 2 ("Cat") based on Stimulus 2 without listening to it (telling mode); (4) Orally answering comprehension questions about Story 2.[6] There was no time limit for completing the tasks, but participants typically took between 30 and 50 minutes to produce both narratives.

---

[6] The present preliminary study focuses only on the written narratives (tasks 1 and 3); oral comprehension questions (tasks 2 and 4) are not analysed here and will be addressed in future research.

3. Data analysis

After collecting the data, the narratives were anonymized, with each participant assigned a numerical and letter code, and then transcribed into a computer-readable format. The analysis of lexical errors was conducted within the framework of EA (Corder, 1974; Ellis, 1994), following the hierarchical structure of error classification (Hoffmann, 2022).

Since the examination of specific types of lexical errors (e.g. the opposition between intralingual and transfer errors, followed by distinctions such as dictionary errors, borrowings, coinages, etc.) was not the focus of this study and constitutes a topic for separate in-depth research, all errors were annotated as lexical. The only exception was spelling errors, which were classified separately and included orthographic and phonological errors.

Each error, regardless of its 'weight', was assigned one point, meaning that errors were not classified as communicatively significant or insignificant, as suggested in the tradition of teaching Russian as a foreign language (e.g. Zhorova, 2005). Repeated errors were counted as many times as they appeared for the following reasons: 1) Some participants, after initially making the same mistake twice, corrected themselves in the middle of the narrative and subsequently wrote the word correctly. 2) Since lexical accuracy was calculated using the formula 'the ratio of lexical errors to the total number of lexical words' (e.g. Leńko-Szymańska, 2019, 107), this counting method was the only appropriate option. Otherwise, the calculation would have involved counting error types in relation to the number of types rather than the actual lexical words.

For counting the number of lexical words in each text, the following approach was used. In this study, functional words were strictly defined as simple, complex, and compound prepositions, conjunctions, and particles (e.g. *и* 'and', *или* 'or', *поэтому* 'so', *потому что* 'because'). However, derivative prepositions (e.g. *вследствие* 'as a result of', *благодаря* 'thanks to') were classified as lexical words. This classification is based on two considerations. First, the use of such prepositions indicates a broader lexical repertoire. Second, certain types of errors occurring in these prepositions are lexical rather than grammatical (e.g. *Мальчик не замечает, что у него больше нет рыб благодаря* кошке.* 'The boy does not notice that he has no more fish thanks to* the cat.'). As a result, they are considered lexical words in accordance with the applied formula.

4. Results

The information about the number of participants in each group, mean age, mean age of acquisition (AoA) for German and Russian, mean LP test score and standard deviation (SD) is summarized in Table 1.

LP test results in HLLs and FLLs						
LL type	Sample size	Mean age	Mean AoA (German)	Mean AoA (Russian)	Mean score in LP test	SD (score in LP test)
HLLs	14	23;14	2;5	0;0	88.79%	5.74
FLLs	11	23;73	0;0	16;18	84.79%	7.05

Table 1: Groups of participants and LP test results in each group

The analyses are performed using R Statistical Software (v4.4.2; R Core Team 2024). The results will be discussed below in relation to each research question.

4.1 Frequency of lexical and spelling errors within the task mode (telling vs. retelling) and across the LL groups (FLLs vs. HLLs)

In Table 2, the mean and SD values of lexical and spelling errors in different task modes (telling and retelling) in FLLs and HLLs along with effect sizes (FLLs vs. HLLs) are presented. FLLs made considerably more lexical errors than HLLs (Cohen's d ≈ 1.70 in telling and ≈ 1.41 in retelling). At the same time, HLLs made considerably more spelling errors than FLLs (Cohen's d ≈ -1.17 in telling and ≈ -1.31 in retelling). According to Cohen (1988), such effect sizes[7] represent meaningful differences between groups.

Error rates (mean, SD) across task modes, and Effect sizes for FLLs and HLLs						
Type of errors	**Task mode**	**Cohen's D**	**Mean (FLLs)**	**SD (FLLs)**	**Mean (HLLs)**	**SD (HLLs)**
Lexical	Telling	≈ 1.70	12.34	6.20	3.86	3.34
Lexical	Retelling	≈ 1.41	7.73	2.60	3.54	3.29
Spelling	Telling	≈ -1.17	4.67	5.25	17.92	15.18
Spelling	Retelling	≈ -1.31	4.92	4.00	18.49	14.13

Table 2: Mean and SD values of errors in telling and retelling along with effect sizes (FLLs vs. HLLs)

To determine whether HLLs and FLLs differ in the frequency of lexical and spelling errors in telling and retelling modes, Welch two sample t-tests were conducted.[8] The results show the following differences.

Differences between FLLs and HLLs in lexical and spelling errors frequency in each task mode								
Type of errors	**Task mode**	**t-statistic**	**df**	**p-value**	**Bonferroni-corrected p (p_{corr}), x4**	**95% confidence interval**	**Mean (FLLs)**	**Mean (HLLs)**
Lexical	telling	4.093	14.492	0.00102*	0.00408*	0.04051018 0.12911125	0.1234	0.0385
Spelling	telling	-3.042	16.757	0.00744*	0.02976*	-0.22448041 -0.04051933	0.0467	0.1792
Lexical	retelling	3.5559	22.993	0.0017*	0.0068*	0.01752320 0.06627174	0.0773	0.0354
Spelling	retelling	-3.4238	15.57	0.00359*	0.01436*	-0.21994237 -0.05149789	0.0492	0.1849

Table 3: Differences between HLLs and FLLs in lexical and spelling errors frequency in each task mode
* significant (p < 0.05, p_{corr} < 0.05), ** not significant (p > 0.05, p_{corr} > 0.05)

---

[7] According to Cohen (1988), effect sizes of d ≥ 0.8 are considered large.
[8] Welch t-test is more reliable when the groups have unequal variances and/or sample sizes.

The analysis shows significant differences in lexical and spelling error frequencies between HLLs and FLLs in both narrative modes. FLLs produce significantly more lexical errors than HLLs in both telling (t = 4.093, p_{corr} < 0.05) and retelling (t = -3.5559, p_{corr} < 0.05). Together with the observed large effect sizes, these results support the idea that HLLs tend to produce fewer lexical errors than FLLs, likely due to their broader implicit lexical and collocational knowledge.

On the contrary, HLLs demonstrate a significantly higher frequency of spelling errors in comparison to FLLs, both in telling (t = -3.042, p_{corr} < 0.05) and retelling (t = 3.4238, p_{corr} < 0.05). A possible explanation for this, as Brüggemann (2018) notes, is that due to the quantitative and qualitative reduction of unstressed vowels, as well as phonetic combinatory processes (such as assimilation, dissimilation, accommodation, and devoicing), Russian exhibits a largely opaque phoneme-grapheme correspondence. Therefore, mastery of Russian orthography relies on explicit knowledge of these processes. This knowledge typically lacks in HLLs as they acquire the language through naturalistic exposure rather than formal instruction. Moreover, HLLs experience a process of palatalization loss (for German Russian – *ibid*, for American Russian – Bermel & Kagan, 2000) which further contributes to the occurrence of spelling errors.

4.2 Correlation between lexical and spelling accuracy in narratives and LL's language proficiency test scores

To examine the relationship between lexical and spelling accuracy in narratives produced by HLLs and FLLs and their language proficiency (anchor test scores), Pearson's correlation analysis was conducted.

Correlation between the lexical and spelling accuracy in different task modes in FLLs and HLLs and their language proficiency test's score					
Type of errors	**LL type**	**Task mode**	**Correlation with LP score (r)**	**p-value**	**Bonferroni-corrected p (p_{corr}), x8**
Lexical	FLLs	Telling	-0.279	0.407**	1.0**
Spelling	FLLs	Telling	0.349	0.292**	1.0**
Lexical	HLLs	Telling	-0.720	0.004*	0.032*
Spelling	HLLs	Telling	-0.321	0.263**	1.0**
Lexical	FLLs	Retelling	-0.532	0.092**	0.736**
Spelling	FLLs	Retelling	0.373	0.258**	1.0**
Lexical	HLLs	Retelling	-0.638	0.014*	0.112**
Spelling	HLLs	Retelling	-0.211	0.469**	1.0**

Table 4: Correlation between the lexical and spelling accuracy in different task modes (telling and retelling) in FLLs and HLLs and their language proficiency test's score
* significant (p < 0.05, p_{corr} < 0.05), ** not significant (p > 0.05, p_{corr} > 0.05)

The results show various patterns between the two groups of learners and across different task modes. For HLLs, a significant negative correlation was found between LP score and lexical error frequency in telling (r = -0.72, p_{corr} = 0.032). In retelling, the correlation was not significant after applying the Bonferroni correction (p_{corr} = 0.112). However, a numerical trend can be observed,

DiSlaw

Didaktik slawischer Sprachen

therefore, further research with a larger sample size is needed.[9] At present, higher general language proficiency in HLLs is strongly associated with a lower frequency of lexical errors in telling. These findings are indirectly in line with previous studies that showed correlations between lexical proficiency and some structural domains. For example, Polinsky (1997, 2007) found that Russian heritage speakers with higher vocabulary proficiency measured through a basic list of 200 words, demonstrated better control of agreement, case marking, and subordination in spontaneous speech.

At the same time, no significant correlation was observed between LP and spelling errors frequency in either telling or retelling. This suggests that HLLs may encounter difficulties with spelling even when their general language proficiency is high. However, further research with a larger sample is needed to confirm this.

For FLLs, no significant correlations were found between lexical errors and LP scores in either telling (r = -0.279, p_{corr} > 1.0) or retelling (r = -0.532, p_{corr} = 0.736). Similarly, no significant correlations were observed between spelling errors and LP scores in either telling (r = 0.349, p_{corr} > 1.0) or retelling (r = 0.373, p_{corr} > 1.0).

These findings would be particularly interesting to compare with the results of a similar study by Llach (2007), which examined the correlation between lexical errors in texts produced by young learners of English as a foreign language (EFL) and two measures of language proficiency: a cloze test and a reading comprehension test. Llach found that lexical error frequency was not related to the cloze test results, but there was a statistically significant relationship between lexical errors and reading comprehension scores.[10] The author concluded that these two tests might assess different aspects of language competence: the cloze test measures discrete-point language knowledge, while the reading comprehension test captures a more general language ability. Applying these findings to the present study and considering that the multiple-choice test used was constructed based on a cloze test (Luchkina et al., 2021), it is particularly interesting that unlike in Llach's study (2007), the HLL group still demonstrated a significant negative correlation between their test results and lexical error frequency in telling. On the one hand, this may suggest that the relationship between different components of language competence in HLLs and FLLs may differ significantly and be shaped by various patterns. On another hand, the results may diverge from Llach's (2007) study on EFL learners due to differences in LP tests construct. Moreover, this might also be explained by a greater degree of lexical similarity between the L1 (Spanish) and the TL (English) in Llach's (2007) study compared to the present study (German vs. Russian). Finally, the discrepancies may stem from the different number of respondents; therefore, further research with a larger sample will be needed to determine this clearly.

---

[9] For a correlation analysis with a medium effect size (r = 0.3) around 88 participants for 80% power and α = 0.05 would be needed. However, the observed large effects (r = -0.720) for HLL's lexical accuracy in telling suggest robust tendency even with a smaller sample.

[10] In the present preliminary study, comprehension questions were not addressed. However, they will be the focus of future research with a larger sample.

5.   Discussion with take away messages for teaching practice

Despite being preliminary, these findings offer valuable insights for the field of FL and HL didactics. Significant differences in lexical and spelling error frequencies between HLLs and FLLs observed in both narrative modes and accompanied by a large effect size, statistically highlight the differences between HLLs and FLLs and further emphasize the need for individualized instructional strategies for different types of LLs. Thus, the high frequency of lexical errors in narratives produced by the FLL participants supports the idea that learners may benefit from structured exposure to frequent and natural collocations and (semi-)fixed expressions. This conclusion aligns with the principles of the Lexical Approach (Lewis, 1993), which views ready-made lexical chunks to be central to language competence. Notably, this approach suggests that LLs should engage with extensive reading and listening materials, focusing on variety rather than depth (Lewis, 1997), which exposes learners to a wide range of lexical chunks and patterns.

In contrast, the HLLs' strong lexical skills suggest the presence of a well-established internalized lexical store, yet their spelling issues point to the need for explicit focus on the orthographic rules of the TL and the written form of words, integrating spelling and morphological awareness into lexical instruction. This is also consistent with findings from previous studies, e.g. that FLLs "do not have the same spelling problems and do not profit from extensive and repetitive spelling assignments that are much needed by heritage learners" (Kagan & Dillon, 2001, 513–514).

A significant negative correlation found between lexical errors frequency and LP score in telling in HLLs, on the one hand, may suggest that for the HLL group, using the result of the multiple-choice test to assess LP may indirectly capture a broader range of linguistic competencies. This may make the use of multiple-choice tests for assessing LP in HLLs more effective than in FLLs. On the other hand, these preliminary findings may further support the idea that lexical accuracy is not just a subskill but a potential predictor of general language proficiency, particularly in HLLs, and highlight the pedagogical relevance of a lexically driven curriculum designed in accordance with LLs' background.

6.   Conclusion with comment on limitations

The study presents preliminary results on the relationship between the frequency of lexical and spelling errors in written elicited narratives of FLLs and HLLs in two elicitation modes (telling and retelling) and the correlation of lexical and spelling errors frequency in HLLs and FLLs with their language proficiency test scores.

The limitations of the present study include, first, the small number of participants. A small sample size always requires more cautious interpretation, since there is a risk that the result may not be as stable as in the case of larger sample sizes. Although even such a small sample was able to reveal several interesting tendencies, they need to be validated by further research with larger sample sizes. Moreover, in accordance with the future research plans, the results of this study will not only be compared with the findings from a larger sample of participants, but will also involve examining other lexical parameters in the texts of FLLs and HLLs, such as lexical diversity, density, and sophistication, which, along with accuracy, are part of the lexical richness model (Read, 2000). Finally, the focus of the study was only on the narratives of participants

with relatively high LP levels, while it would be interesting to explore whether the identified trends also apply to participants at lower proficiency levels.

References

Beaudrie, S. M. (2011). *Spanish heritage language programs: A snapshot of current programs in the southwestern United States. Foreign Language Annals, 44*(2), 321–337. https://doi.org/10.1111/j.1944-9720.2011.01137.x

Berkoff, N. A. (1982). Error Analysis Revisited. In G. Nickels & D. Nehls (eds.), *Error analysis, contrastive linguistics, and second language learning: Papers from the 9ᵗʰ International Congress of Applied Linguistics.* J. Groos.

Bermel, N., & Kagan, O. (2000). The maintenance of written Russian in heritage speakers. In O. Kagan & B. Rifkin (eds.), *The learning and teaching of Slavic languages and cultures* (405–436). Slavica Publishers.

Benmamoun, E., Montrul, S., & Polinsky, M. (2013). Heritage languages and their speakers: Opportunities and challenges for linguistics. *Theoretical Linguistics, 39*(3–4), 129–181. https://doi.org/10.1515/tl-2013-0009

Brüggemann, N. (2018). SchreibanfängerInnen im herkunftssprachlichen Russischunterricht. In A. Bergman, O. Caspers, & W. Stadler (eds.), *Didaktik der slawischen Sprachen – Beiträge zum 1. Arbeitskreis in Berlin (12.–14.09.2016)* (177–192). Innsbruck University Press.

Cohen J. (1988). *Statistical power analysis for the behavioral sciences*. 2ⁿᵈ ed. Lawrence Erlbaum Associates.

Corder, S. P. (1974). Error analysis. In J. P. B. Allen & S. P. Corder (eds.), *Techniques in applied linguistics*. Vol. 3. (122–154). Oxford University Press.

Dubinina, I., Mustajoki, A. S., Protasova, E. I., & Vakhtin, N. B. (2010). How to ask for a favor: A pilot study in heritage Russian pragmatics. In A. S. Mustajoki, E. I. Protasova, & N. B. Vakhtin (eds.), *Instrumentarium of linguistics: Sociolinguistic approaches to non-standard Russian* (418–431). Department of Modern Languages, University of Helsinki.

Ellis, R. (1994). *The study of second language acquisition*. Oxford University Press.

Ermakova, N. (2023). *Heritage speakers in Russian teacher training programs: A comparison of curricular frameworks for enhancing linguistic and communicative competences of future teachers in Germany. DiSlaw – Didaktik slawischer Sprachen, 1*(1), 1–13. https://doi.org/10.48789/2023.1.2

Flecken, M. (2011). Assessing bilingual attainment: Macrostructural planning in narratives. *International Journal of Bilingualism, 15*(2), 164–186. https://doi.org/10.1177/1367006910381187

Hemchua, S., & Schmitt, N. (2006). An analysis of lexical errors in the English compositions of Thai learners. *Prospect, 21*(3), 3–25.

Housen, A., & Kuiken, F. (2009). Complexity, accuracy, and fluency in second language acquisition. *Applied Linguistics, 30*(4), 461–473. https://doi.org/10.1093/applin/amp048

Gagarina, N., Klop, D., Kunnari, S., Tantele, K., Välimaa, T., Balčiūnienė, I., Bohnacker, U., & Walters, J. (2012). MAIN: Multilingual Assessment Instrument for Narratives. *ZAS Papers in Linguistics, 56*, 155–170. https://doi.org/10.21248/zaspil.56.2019.414

Gagarina, N., Klop, D., Kunnari, S., Tantele, K., Välimaa, T., Bohnacker, U., & Walters, J. (2019a). MAIN: Multilingual Assessment Instrument for Narratives – Revised*. ZAS Papers in Linguistics, 63*, 1–20. https://doi.org/10.21248/zaspil.63.2019.516

Gagarina, N., Bohnacker, U., & Lindgren, J. (2019b). Macrostructural organization of adults' oral narrative texts. *ZAS Papers in Linguistics, 62*, 190–208. https://doi.org/10.21248/zaspil.62.2019.449

Gutiérrez-Clellen, V. F., Restrepo, M. A., Bedore, L., Peña, E., & Anderson, R. (2000). Language sample analysis in Spanish-speaking children: Methodological considerations. *Language, Speech, and Hearing Services in Schools, 31*(1), 88–98. https://doi.org/10.1044/0161-1461.3101.88

Heilmann, J., Miller, J. F., Nockerts, A., & Dunaway, C. (2010). Properties of narrative retells: A clinical tool for assessing school-age children's language development. *Language, Speech, and Hearing Services in Schools, 41*(2), 204–219. https://doi.org/10.1044/1058-0360(2009/08-0024)

Hoffmann, T. (2022). Measuring lexical accuracy: The categorization of lexical errors in written L2 English. In A. Leńko-Szymańska & S. Götz (eds.), *Complexity, accuracy and fluency in learner corpus research* (159–179). John Benjamins. https://doi.org/10.1075/scl.104.07hof

Jia, G., & Aaronson, D. (2003). A longitudinal study of Chinese children and adolescents learning English in the United States. *Applied Psycholinguistics, 24*(1), 131–161. https://doi.org/10.1017/S0142716403000079

Kagan, O., & Dillon, K. (2001). A New Perspective on Teaching Russian: Focus on the Heritage Learner. *The Slavic and East European Journal*, *45*(3), 507–518. https://doi.org/10.2307/3086367

Karlsen, J., Geva, E., & Lyster, S. (2016). Cognitive, linguistic, and contextual factors in Norwegian second language learner's narrative production. *Applied Psycholinguistics, 37*(5), 1117–1145. https://doi.org/10.1017/S014271641500051X

Krasnoshchekova, S., & Kashleva, K. (2019). Narrative Competence of Adult L2 Russian Learners. *Journal of psycholinguistic research, 48*(3), 617–641. https://doi.org/10.1007/s10936-018-9622-3

DiSlaw
Didaktik slawischer Sprachen

Kupisch, T., Barton, D., Hailer, K., Klaschik, E., Stangen, I., & Lein, T. (2014). Foreign accent in adult simultaneous bilinguals. *Heritage Language Journal, 11*(2), 123–150. https://doi.org/10.46538/hlj.11.2.2

Leńko-Szymańska, A. (2019). *Defining and Assessing Lexical Proficiency.* Routledge. https://doi.org/10.4324/9780429321993

Lewis, M. (1993). *The lexical approach: The state of ELT and a way forward*. Language Teaching Publications.

Lewis, M. (1997). *Implementing the lexical approach: Putting theory into practice*. Language Teaching Publications.

Llach, M. P. A. (2007). Lexical errors as writing quality predictors. *Studia Linguistica*, *61*(1), 1–19. https://doi.org/10.1111/j.1467-9582.2007.00127.x

Luchkina, T., Ionin, T., Lysenko, N., Stoops, A., & Suvorkina, N. (2021). Evaluating the Russian Language Proficiency of Bilingual and Second Language Learners of Russian*. Languages, 6*, 83. https://doi.org/10.3390/languages6020083

Marian, V., Blumenfeld, H. K., & Kaushanskaya, M. (2007). The Language Experience and Proficiency Questionnaire (LEAP-Q): assessing language profiles in bilinguals and multilinguals. *Journal of speech, language, and hearing research (JSLHR)*, *50*(4), 940–967. https://doi.org/10.1044/1092-4388(2007/067)

Mayer, M. (1969). *Frog, where are you?* Dial Books for Young Readers.

Montrul, S. (2004). *The acquisition of Spanish: Morphosyntactic development in monolingual and bilingual L1 acquisition and in adult L2 acquisition*. John Benjamins.

Montrul, S. (2008). *Incomplete Acquisition in Bilingualism: Re-examining the Age Factor*. John Benjamins. https://doi.org/10.1075/sibil.39

Montrul, S. (2012). Is the heritage language like a second language? *EUROSLA Yearbook, 12*(1), 1–29. https://doi.org/10.1075/eurosla.12.03mon

Montrul, S. (2016). *The acquisition of heritage languages*. Cambridge University Press.

Montrul, S. (2023). Heritage Languages: Language Acquired, Language Lost, Language Regained. *Annual Review of Linguistics*, *9*, 399–418. https://doi.org/10.1146/annurev-linguistics-030521-050236

O'Grady, W., Archibald, J., Aronoff, M., & Rees-Miller, J. (2001). *Contemporary linguistics: An introduction.* 4th ed. St. Martin's Press.

Pavlenko, A. (2008). Emotion and emotion-laden words in the bilingual lexicon. *Bilingualism: Language and Cognition, 11*(2), 147–164. https://doi.org/10.1017/S1366728908003283

Pearson, P. D. (2002). The effects of phonological awareness and letter knowledge on early literacy skills. *Reading Research Quarterly, 37*(3), 382–394. https://doi.org/10.1598/RRQ.37.3.5

Polinsky, M. (1995). American Russian: Language loss meets language acquisition. In W. Browne, E. Dornisch, N. Kondrashova, & D. Zec (eds.), *Formal approaches to Slavic linguistics: The Cornell meeting* (370–406). Michigan Slavic Publications.

Polinsky, M. (2006a). Incomplete acquisition: American Russian*. Journal of Slavic Linguistics, 14,* 161–219.

Polinsky, M. (2006b). *Heritage language narratives. Heritage Language Education: A New Field Emerging* (108–156). Routledge.

Polinsky, M. (2008). Heritage language narratives and incomplete acquisition. *Heritage Language Journal, 6(1),* 40–71.

Polinsky, M. (2010). A composite linguistic profile of a heritage speaker of Russian. *The Learning and Teaching of Slavic Languages and Cultures* (437–466). Slavica Publishers.

Polinsky, M. (2015). Heritage languages and their speakers. *Theoretical Linguistics, 39*(3–4), 129–181.

R Core Team (2024). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria.

Read, J. (2000). *Assessing vocabulary.* Cambridge University Press.

Sekerina, I. A. (2013). A psychometric approach to heritage language studies. *Heritage Language Journal, 10*(2), 1–12. https://www.researchgate.net/publication/260087928_2013_A_Psychometric_Approach_to_Heritage_Language_Studies

Schneider, P., Hayward, D., & Dubé, R. V. (2006). Storytelling from pictures using the Edmonton narrative norms instrument. *Canadian Journal of Speech-Language Pathology and Audiology, 30*(4), 224–238.

Skehan, P. (1988). *The role of strategy in second language learning. Australian Review of Applied Linguistics, 11*(1), 93–108.

Sorace, A., & Serratrice, L. (2009). Internal and external interfaces in bilingual language development: Beyond structural overlap. *International Journal of Bilingualism, 13*(2), 195–210. https://doi.org/10.1177/1367006909339810

ZAS = https://main.leibniz-zas.de/en/main-materials/main-materials/ (27.02.2025)

Zhorova, A. P. (2005). = Жорова А. П. (2005). *Коммуникативно значимые ошибки, допускаемые при выполнении субтеста «Письмо» первого уровня // Тестовые формы контроля по русскому языку какиностранному (РКИ).* ЦМО МГУ им. М. В. Ломоносова.