Russian-as-a-heritage-language vocabulary acquisition by bi-/multilingual children in Canada

Veronika Makarova¹, Natalia Terekhova²

¹University of Saskatchewan
914 9 Campus Dr, Arts Building, Saskatoon, S7N 5A5, Saskatchewan, Canada
²Options Community Services
9815 140th St, Surrey, V3T 4M4, British Columbia, Canada

Abstract. The significance of this paper is in its contribution to the innovative and rapidly developing research area of Russian as a heritage language (RHL) around the world. The purpose of the reported study is to explore Russian vocabulary development by bi-/multilingual children acquiring Russian as a heritage language in Canada. The materials come from vocabulary development and non-canonical lexical forms (NCF, earlier known as “errors”) in the speech of 29 bi-/multilingual children (between the ages of 5 and 6) from immigrant families in Saskatchewan, Canada (RHL group) as well as of 13 monolinguals from Russia (MR group). The study employs a method of a comparative analysis of vocabulary in picture-prompted narratives by children from the above two groups. The results demonstrate that bi-/multilingual RHL speaking children produced significantly more lexical NCFs as compared to their monolingual peers (MR), whereas narrative length in words, speech rate in wpm and vocabulary size did not differ across the two groups. Most NCFs in the RHL sample related to the use of verbs, followed by NCFs in the use of nouns. Unlike the speech of MR speakers, RHL participants’ language use exhibits some slight impact of dialectal forms, a few borrowings from English and code-switches to English. The study has applications for the theory of bi-/multilingualism as well as for teaching RHL to children of immigrants in North American and other contexts.

Keywords: child bilingualism and multilingualism, lexis, Russian language abroad, families of Russian immigrants in Canada, Russian as a heritage language

Article history: received: 10.05.2020; accepted: 22.07.2020.

Acknowledgments: The authors express their gratitude to all participants of the study and their families.

Introduction

The studies of speech production by heritage language speakers inform the linguistic theory about the fundamental processes in language acquisition and language structure (Scontras et al., 2015), since heritage languages are “a particular phenomenon within bilingualism” (Polinsky, 2018a: 547). Heritage languages are first (native) or home languages acquired by bi-/multilingual children in the environments where another language is dominant in society (Lorenz et al., 2019; Polinsky, 2018b). The majority language is often acquired early from the age of between three and six (Lorenz et al., 2019), i.e., heritage language speakers become bilingual (or multilingual) in their heritage language, the dominant language of the society and possibly in another language.

The competency of heritage speakers in their mother tongue strongly varies by the speaker and the circumstances of language acquisition (Kupish, Rothman, 2016; Polinsky, Kagan, 2007). They could become balanced bilinguals and develop high competency in both the majority and heritage language, or the heritage language skills could decline with time (Flynn et al., 2005; Polinsky, 2011; 2018b). Multiple factors are engaged in the heritage vs. majority language development including the speakers’ age, the amount and quality of language exposure, proximity of the languages involved, age of the child upon immigration to the host country, family language policies and practice, etc. (Fernandes, 2019; Lorenz et al., 2019; Smyslova, 2012). Many earlier studies describe issues in heritage language speaker’s competencies in a comparison to the language baseline presented by their monolingual peers (Bennamoun, 2013; Polinsky, 2018b; Polinsky, Kagan, 2007). Competences in the heritage language often remain strong in comprehension and production, but may be low in reading and writing skills (Montrul, 2011; 2015), since the development of the latter two competencies is typically associated with literacy skills obtained through formal education which is often not available or limited in a heritage language (Montrul, 2015).

Vocabulary has been identified in earlier studies as one of the major “deficits” of heritage language speakers in a number of languages (e.g., Scontras et al., 2015). Some differences were observed in vocabulary development and lexical retrieval between heritage and monolingual speakers (e.g., Jia, Paradis, 2015; Silvén et al., 2014; Yan, Nicoladis, 2009).

Not very many studies of child Russian as a heritage language (RHL) speakers are available worldwide (e.g., Bar-Shalom, Zaretsky, 2008; Klassert et al., 2012), but some studies did show the divergence in vocabulary acquisition by RHL child bilinguals as compared to Russian monolingual children (MR) (Ringblom, Dobrova, 2019), as well as the occurrence of lexical errors in RHL speech production (e.g., Polinsky, 2005; Bar-Shalom, Zaretsky, 2008).

There is a gap in research data related to RHL studies in Canadian context. Overall, RHL studies in Canada are much less developed than in the US, where an the interest in Russian-English bilingualism and heritage Russian was largely triggered by growth in the number of Russian-speaking immigrants and an increased number of second-generation immigrants with varied levels of Russian proficiency signing up to take university-level Russian language classes in the early 21st century (Makarova, 2012: XIII). While Canada has close to 200,000 individuals
speaking Russian as a mother tongue, not many universities offer regular Russian as a foreign language courses, and very few – Russian-as-a-heritage-language courses (Makarova, 2020). Consequently, only very few studies address Russian-English bilingualism in the country (e.g., Kazanina, Phillips, 2007; Nicoladis et al., 2016; Makarova et al., 2017), and none of them focus on non-canonical lexical forms in RHL speech.

While earlier research referred to “deficiencies” or “errors” in linguistic outputs by bi-/multilingual children (e.g., Bar-Shalom, Zaretsky, 2008; Benmamoun, 2013), we use the term “non-canonical” forms to denote speech constituents that differ from standard adult language use (e.g., Antomo, Müller, 2018: 5), since we do not see bi-/multilingualism as a “deficiency”, and the “errors” often originate in the developmental processes of language acquisition that can be similar among bilinguals and multilinguals (Makarova, 2020).

The study reported in this article considers vocabulary development and non-canonical lexical forms (NCF) in the speech of bi-/multilingual children acquiring Russian as a heritage language in Saskatchewan, Canada. Russian as a heritage language in Saskatchewan has a small community of about 1500 Russian speakers. No government or institutional support is available for the language maintenance, and no Russian language courses are available in the province at any level of education (Makarova, 2020). The aims of the study were to investigate how vocabulary acquisition by RHL speaking children (age 5–6 years old) in this environment may be comparable to that of their monolingual peers; to describe the specific features of vocabulary development by RHL speaking children as well as NCFs that may be present in their vocabulary use with reference to the lexical development in the speech of their monolingual peers in Russia.

The research questions of the study were:

1) what is the overall level of lexical development reflected in the speech of child Russian as a heritage language speakers in Saskatchewan, Canada, and how does this level compare to the vocabulary development of their MR peers?

2) what kind of non-canonical lexical forms do child RHL speakers produce in heritage Russian?

**Materials and methods**

Two groups of participants were recruited by means of purposive sampling. The first group included 29 RHL speaking bi-/multilingual children from Saskatoon, Saskatchewan, within the age group between 5 and 6 years old (11 boys and 18 girls). All the RHL group participants were proficient in Russian (acquired at home) and English (acquired outside of home environment). The requirements towards participation were as follows:

– the participating child had to be either born in Canada or brought to the country before the age of 3 years old (to exclude the factors of age upon arrival and attendance of a school in the home country);

– age between 5 and 6 years old (the age when a heritage language has reached a relatively high level of proficiency and remains a mother tongue);

– attendance of a preschool/school in Saskatchewan for at least 10 months (to ensure establishing of bilingualism);
– fluent spoken proficiency in Russian and English self-reported by the child and the child’s parent.

The second group of participants were 13 Russian speaking monolinguals (6 girls and 7 boys) from Kemerovo, Russia, also in the age group between 5 to 6 years old. The speech production by this group (MR sample) was used as a frame of reference to describe the vocabulary use by the first group.

Both groups of participants were requested to tell a story represented in a set of six pictures (from a children’s online picture book, “Dobraya skazka v kartinkax” (“A good fairy-tale in pictures”). The picture-prompted narratives were recorded with a Zoom H2n Handy Recorder in Wave Sound format. The narratives were manually transcribed and subjected to linguistic analysis to examine lexical characteristics of the participants’ Russian speech. One-way Univariate ANOVAs were conducted to compare some vocabulary development parameters across the groups.

This article provides some quantitative comparisons of the lexical parameters in the speech of the two participant groups and focuses on the qualitative descriptions of the bi-/multilinguals’ vocabulary and the types of NCFs observed in their lexical use.

Results

RHL and MR speech samples: some characteristics. The following general parameters related to lexical development have been extracted from the data: narrative length in words (i.e. the total number of words used in a child’s narrative), the number of different vocabulary tokens (number of different lexemes that occurred in the narrative, excluding the repetitions of the same word/word-form), the number of words per utterance, speech rate (in number of words per minute), and the number of lexical NCFs. The parameter values across the two participant groups are represented in Figures 1 and 2.

The parameter of narrative length in words ($df = 41; F = 0.28; p = 0.59$) is slightly higher in the RHL group ($x = 175.0; SD = 91.2$) than in the MR group ($x = 141.3; SD = 76.2$), but there is no significant difference between the two groups (Figure 1).

The number of different vocabulary tokens or lexemes ($df = 41; F = 0.0002; p = 0.99$) is about the same in both groups (RHL $x = 67.3; SD = 29.6$; MR $x = 67.2; SD = 25.2$) (Figure 1).

The number of words per utterance ($df = 41; F = 0.139; p = 0.71$) is insignificantly higher in the RHL group ($x = 7.6; SD = 6.22$) than in the MR group ($x = 7.0; SD = 4.1$) (Figure 1).

The speech rate is higher in the RHL group ($x = 76.8; SD = 24.2$) than in the MR group ($x = 70.9; SD = 27.3$), but this difference is not significant for the given samples ($df = 41; F = 0.49; p = 0.48$) (Figure 1).

The number of lexical errors is significantly higher ($df = 41; F = 6.8; p = 0.01$) in the RHL group ($x = 1.5; SD = 0.22$) than in the MR group ($x = 0.5; SD = 0.8$) (Figure 2).

The narratives by the children in both groups were short, and the vocabulary was repetitive (about 67 different vocabulary tokens per 150 words total in a nar-
due to the nature of the task: words related to main characters (animals) in the pictures, personal pronouns, and verbs of motion were repeated multiple times. The sample is therefore not sufficient for building a comprehensive picture of children’s language development, but it does help to provide some insights into the process.

**Figure 1. Some vocabulary parameters in RHL and MR samples**

**Lexical NCFs in the RHL sample.** The RHL speech samples contained a total of 28 lexical NCFs, i.e. the use of words unsuitable to the given context, or the use of non-existing words (occasionalisms). 15 of these NCFs were in verbs, 12 were in nouns, and 1 was in adverb use.

Lexical NCFs in the use of verbs in the RHL sample were typically associated with a more generic verb employed instead of a more specific one that the chil-
Children likely did not know or remember (Examples 1 and 2). Some of these NCFs occurred in cases in which the verbal use was idiomatic (Example 4). The children also showed signs of incomplete word acquisition, producing verbal forms that resemble the standard form, but which are malformed (occasionalisms). In particular, the verb “to ignite” caused difficulties for three children (Examples 4 and 5). In two cases, the verbal forms had distorted sound structures (Examples 6 and 7).

Noun errors by RHL children also reveal similar tendencies, whereby a more generic noun is substituted for a more specific one (Example 8). In particular, two participants substituted the word “nora” (a burrow) for the more generic “dyrka” (hole) (Example 9), and one participant – for “kamenička” (an occasionalism likely derived from the dialectal “kamnica” (a stone construction)). In one case, a participant used the word “kukareku” (a sound produced by rooster, i.e. “cockadoodledo”) instead of the noun for “rooster” (petux), a word that the child did not know or forgot. One participant used the word “čikin” (chicken, a borrowing from English that immigrants from Russia sometimes use in their Russian speech) instead of the Russian word for “hen” (kurica). Two participants confused the word “šaški” (checkers) with similar sounding words with different meanings: “šajki” (buckets) and “šalaški” (sheds). One child was likely misled by the words “pčёlkа” (little bee) and “bloška” (little flea) and produced a blend word form “ploška” instead (which in adult language means “a bowl”). The NCF in adverb use is provided below in Example 10, in which a child employed the adverb “hard” instead of “fast” in a context requiring the phrase “ran fast.”

Example 1
(1) Speaker RHL 1: – i oni vverx na nёm xodili
and they went up on it
Standard: – i oni naverx na nёm (po)pily
and they swam up on it

The Russian standard requires in this context the use of the verb “swim” and not “go.”

Example 2
(2) Speaker RHL 9: – oni delali čaj
they did* tea

* – the translation attempts to render the erroneous lexical use in English.

Standard: – oni zavarili/pili čaj
they brewed/made tea

Example 3
(3) Speaker RHL 19: – oni sygrali večerinku
they played a party

Standard: – oni ustroiši večerinku
they organized a party

The verb “sygrat’” (play) in Russian is used with the word “wedding” and words denoting games, but not with the word “party”.
Example 4
(4) Speaker RHL 2: – oni razžixli kastёр
– they ignited [occasionalism] fire
Standard: – oni razojgli kastёр
– they ignited the fire

Example 5
(5) Speaker RHL 16: – oni zagali ogon’
– they ignited [occasionalism] fire
Standard: – oni zažgli ogon’
– they ignited the fire

Example 6
(6) Speaker RHL 14: – lisa udušala
– fox [not clear], possibly “ubežala” (ran away)

Example 7
(7) Speaker RHL 18: – oni dubačili lisu
– they beat up [distorted sound form] [the] fox
Standard: – oni dubasili lisu
– they beat up the fox

Example 8
(8) Speaker RHL 26: – a zajčik v ruke lisicy
– and [the] bunny is in [the] hand of the fox
Standard: – a zajčik v lapi lisicy
– and [the] bunny is in [the] paw of the fox

Example 9
(9) Speaker RHL 6: – a lisa v dyrke
– and the fox [is in a] hole
Standard: – a lisa v nore
– and the fox [is in a] burrow

Example 10
(10) Speaker RHL19: – on sil’no bežal
– he ran hard
Standard: – on bystro bežal
– he ran fast

The Russian monolingual sample contained only 5 lexical errors, three in verb use, one in noun use, and one in adverb use, e.g., “igrali v peški” (played pawns) instead of “igrali v šaški” or “šahmaty” (played checkers or chess). Overall, the levels of lexical development across the two groups appear quite similar, but RHL bi-/multilingual speakers made more lexical errors than monolinguals.

In the speech of five participants whose parents came from Ukraine, there were some slight influences of South-Eastern Russian/Ukrainian dialects in the sound constituents of some words. Four participants (RHL 6, 11, 16 and 18 used the /ɦ/ phoneme instead of standard Russian /ɡ/). Participant RHL 22 demonstrated some more dialectal influences by pronouncing the words “përli” and “sidjat” [ˈpʲorlying], [ˈsidjɪt].
[sʲiˈdʲat] with Southern Russian/Ukrainian sound equivalents: “perli” and “sydjat” [ˈpɛrli], [sɨˈdʲat].

Colloquial forms. Ten colloquial pronunciation forms occurred in the speech of 10 RHL children, and 5 in the speech of 2 MR children. For example, participants RHL2 and R10 used the form “čě” in place of the standard “što” (what). Participants RHL3 and RHL11 pronounced the word “net” (no) [nʲet] as “ne-a” [nʲe ḭa], which is typical in colloquial casual speech.


Code-switches in the RHL sample. Only five RHL participants code-switched to English in a total of fifteen instances. Code-switches were primarily used in the narratives of speakers with lower Russian proficiency to substitute English equivalents for the words they did not know in Russian. The remaining 24 participants did not code-switch at all, sticking only to Russian in their narratives.

Speaker RHL 4, who had a lower RHL proficiency, code-switched to English four times, once probably just to show off his English skills (“Oh, super duper magic!”) and three times to indicate his boredom with the picture description task and to hint that he did not want to participate any more: e.g., “I am pooped out,” “Èta že [this is] totally boring!”

Speaker RHL 5 asked in Russian, “What do you call it?” about the word for “tree” in Russian (which he had forgotten), then used the English word “tree,” and finally self-corrected for the Russian equivalent “derevo” as he then remembered the word.

Similarly, speaker RHL 22 used the English words “rabbit” and “chicken” instead of their Russian equivalents, as he did not know or forgot the Russian equivalents.

Speaker RHL 7 used code-switches six times, mostly to substitute English equivalents for the words he did not know in Russian, such as “bike” and “squirrel” (Example 11), and once to ask what a boat was called in Russian. In other cases, when Speaker RHL 7 forgot how to say something in Russian, he asked the researcher in Russian to help out (“What is this called? How do you say this?” (Example 12).

Example 11
(53) Speaker RHL 7: – tut est’ medved’, ego dom i ego bike
– here is [the] bear, his house and his bike

Example 12
(54) Speaker RHL 7: – ona take uot …who is that?
– she takes this …who is that?

Speaker RHL 8 counted picture numbers in English before starting the narrative, but then employed no further code-switches to English in the course of his narrative.

Discussion

Vocabulary size (along with grammar development) is typically used as a major parameter of child language assessment (e.g., Vinarskaja, Bogomazov 2005). While Russian 5-year-olds are generally expected to have a vocabulary of 2000–2500
words (Vinarskaja, Bogomazov 2005), this parameter could not be measured in our study, as the nature of the picture description task only elicited a small part of the children’s vocabulary. The lexical development of bi-/multilingual children in the study was overall on par with the development of their monolingual peers in Russia, since there were no significant differences in the total number of words or a total number of different lexemes produced in the speech samples coming from the two groups. However, the bi-/multilingual children produced significantly more lexical non-canonical forms than their monolingual peers (1.5 per child on the average for the bi-/multilingual group vs. 0.5 errors per child for the monolingual group). The results of the study confirm some earlier findings outlining some lexical difficulties experienced by heritage Russian speakers (Polinsky, 2005).

As in many other studies of child speech development (e.g., Ushakova, 2004; Cejtlin, 2009; Panfilova, 2011; Gleason, Bernstein Ratner, 2013), the participants in our study invented words when they did not know or could not recall the target adult word. Quite a few lexical NCFs were caused by children’s under-acquisition of vocabulary, whereby the NCF forms they produced had a phonetic similarity to the target lexeme. This seems to support the theory of phonetic storage of vocabulary in the mental lexicon, i.e., the storage of the words in the brain based on sound similarities (Hoff, 2014: 137).

Most NCFs were observed in verbs, which confirms findings that demonstrate that heritage speakers have some difficulties producing verbal forms (Bonfatti-Sabioni, 2018)

**Code-switches.** Code-switches are very common among bi-/multilingual children (Hoff, 2014; Genesee, Nicoladis, 2007). While code-switching among adults can be explained by multiple factors (Hoff, 2014), in this study, code-switches were primarily caused by children’s lack of knowledge of a Russian word or not being able to remember it. It is worth noting that only children with lower RHL proficiency used code-switches to English, and that the number of code-switches in the whole corpus was very small (15 code-switches in total). It is possible that the number of code-switches was affected by the interviewer, who was a native speaker of Russian, and by a present Russian-speaking parent. The children might therefore have attempted to “stick to” Russian to accommodate to the interviewer and the parent. It is possible that in communication with RHL peers or siblings the number of code-switches could have been higher.

**The use of dialectal forms.** The specific demographic origins of the child bilinguals’ families in our study (over 50% of participants’ families were from Eastern Ukraine) likely contributed to the Southern Russian/Ukrainian dialectal features in our sample. The results seem to suggest that in the RHL context, due to immigration from different countries and regions, dialectal influences could be overall much stronger than in the monolingual samples.

**The use of colloquial forms.** The use of colloquial and vernacular forms by children has not been given sufficient attention so far, and yet, as our study shows, both bi-/multilingual RHL and monolingual Russian speaking children employ a number of colloquialisms, such as “čë” [tʃo] for “čto” [ʃto]. It would be interesting to determine in future research whether colloquial/vernacular forms are retained in the speech of children as they grow up or are substituted with standard versions.
Conclusion

The study shows some common features in the vocabulary development between bi-/multilingual speakers of heritage Russian and their monolingual Russian speaking peers. These similarities include occasionalisms (or the use of words and word forms invented by children), substitutions of more specific words for more generic ones and the use of colloquial/vernacular forms. These features are also typical for both monolingual and bi-/multilingual children across multiple languages.

Some specific features associated with the development of heritage language in immigrant minority settings were also identified, such as the use of dialectal sound constituents of words and code-switches to English.

The bi-/multilingual participants in our study (aged 5–6) show a level of Russian language acquisition similar to the linguistic development of their monolingual peers in Russia, likely because all the participants had a high level of language exposure in the families. It is possible to expect a slowing down in vocabulary acquisition in the heritage language as the children grow older and are increasingly more exposed to the dominant language in the environment, particularly through schooling.

Formal (in-school) education provides children with an immense language booster. It generates rapid growth in literacy; vocabulary expansion related to many areas of the humanities, sciences, and social sciences; and academic language discourse abilities — and it contributes to development of diverse language styles (Hoff, 2014: 275). Lack of formal schooling in a heritage language in Saskatchewan is likely to be detrimental for a further development of children’s RHL vocabulary and RHL on the whole.

References


Усвоение лексики русского как языка наследия детьми-билингвами и полилингвами в Канаде

В. Макарова1, Н. Терехова2

1Университет Саскачевана
Канада, S7N 5A5, Саскачеван, Саскатун, Arts Building, 9 Campus Dr, 914
2Options Community Services
Канада, V3T 4M4, Британская Колумбия, Суррей, 140th St, 9815

Аннотация. Актуальность статьи обусловлена ее вкладом в инновационную разрастающуюся область исследований русского как языка наследия (РЯН) за рубежом. Цель исследования состоит в изучении усвоения лексики РЯН детьми-билингвами и полилингвами, проживающими в Канаде. Материалами исследования послужили лексика и неканонические лексические формы (НЛФ, ранее называвшиеся ошибками) в речи 29 детей-билингвов/полилингвов в возрасте от 5 до 6 лет, членов семей иммигрантов, проживающих в провинции Саскачеван в Канаде (группа РЯН), а также 13 русскоязычных детей-монолингвов (группа МР) из России. В исследовании используется метод сравнительного анализа лексики на протяжении описаний картинок-иллюстраций детьми из двух вышеуказанных групп. Результаты показали, что в речи детей-билингвов/полилингвов наблюдается большее количество НЛФ по сравнению с группой монолингвов, тогда как длина рассказа, речевой темп и лексический запас сходны между группами. Боль-
шинство НЛФ в подкорпусе РЯН относились к формам глаголов, немного меньше НЛФ наблюдалось в формах существительных. В отличие от группы монолингвов, в речи канадской группы были обнаружены диалектные формы, заимствования из английского и переключение кодов. Перспективы исследования относятся к дальнейшей разработке теории билингвизма/полилингвизма, а также преподаванию РЯН детям иммигрантов в Северной Америке и других регионах мира.

Ключевые слова: дети-билингвы и полилингвы, лексика, русский за рубежом, семьи русскоязычных иммигрантов в Канаде, русский как язык наследия

История статьи: поступила в редакцию: 10.05.2020; принята к печати: 22.07.2020.

Благодарности: Авторы выражают благодарность всем участникам исследования и членам их семей.


Сведения об авторах:
Макарова Вероника, профессор кафедры лингвистики Университета Саскачевана. Работала в университетах Японии, России и Великобритании. Результатом ее исследований стали четыре книги, выпущенные под ее редакцией, одна монография и более 100 опубликованных научных статей. Получила три гранта от Канадского государственного финансового агентства (SSHRC) и несколько международных грантов. Сфера научных интересов: лингвистика, прикладная лингвистика, анализ дискурса, изучение языков наследия, двуязычие/многоязычие, социолингвистика, этнические исследования. E-mail: v.makarova@usask.ca

Терехова Наталья, получила степень бакалавра английского языка и литературы в Кемеровском государственном университете, а затем степень магистра лингвистики в Университете Саскачевана с исследовательской направленностью на поддержание языка у билингвов. По окончании учебы координировала услуги письменно-го и устного перевода для иммигрантов. В настоящее время работает с молодежью в качестве разработчика вакансий в Options Community Services (OCS) и волонтером в Исполнительном комитете Канадского совета переводчиков, терминологов и переводчиков. Сфера научных интересов: билингвизм, мультилингвизм, перевод. E-mail: natalia.terekhova@usask.ca