



Cadernos de Linguística  
ISSN:  
ISSN: 2675-4916  
miguel@fale.ufal.br  
Associação Brasileira de Linguística  
Brasil

# BRAPORUS, CORPUS DE FALA COLETADA EM FALANTES DE RUSSO DE HERANÇA NO BRASIL: PROTOCOLO DE COLETA DE DADOS

**SMIRNOVA HENRIQUES, Anna**  
**SKOROBOGATOVA, Aleksandra S.**  
**KACHKOVSKAIA, Tatiana V.**  
**SKRELIN, Pavel A.**  
**RUSEISHVILI, Svetlana**  
**MADUREIRA, Sandra**  
**SEKERINA, Irina A.**

BRAPORUS, CORPUS DE FALA COLETADA EM FALANTES DE RUSSO DE HERANÇA NO BRASIL:  
PROTOCOLO DE COLETA DE DADOS

Cadernos de Linguística, vol. 3, núm. 1, e629, 2022

Associação Brasileira de Linguística

Disponível em: <https://www.redalyc.org/articulo.oa?id=689174650005>

DOI: <https://doi.org/10.25189/2675-4916.2022.v3.n1.id629>



Esta obra está bajo una Licencia Creative Commons Atribución 4.0 Internacional.



REGISTRO DE PROJETO  
BRAPORUS, CORPUS DE FALA COLETADA EM FALANTES DE RUSSO  
DE HERANÇA NO BRASIL: PROTOCOLO DE COLETA DE DADOS

*Anna SMIRNOVA HENRIQUES*

*Pontifícia Universidade Católica de São Paulo (PUC-SP), Brasil*

anna.smirnova.liaac@gmail.com

*Aleksandra S. SKOROBOGATOVA*

*Universidade de São Paulo (USP), Brasil*

as.skorobogatova@gmail.com

*Tatiana V. KACHKOVSKAIA*

*Saint Petersburg State University (SPbU), Federación de Rusia*

tania.kachkovskaya@gmail.com

*Pavel A. SKRELIN*

*Saint Petersburg State University (SPbU), Federación de Rusia*

skrelin@phonetics.pu.ru

*Svetlana RUSEISHVILI*

*Universidade Federal de São Carlos (UFSCar), Brasil*

s.ruseishvili@gmail.com

*Sandra MADUREIRA*

*Pontifícia Universidade Católica de São Paulo (PUC-SP), Brasil*

sandra.madureira.liaac@gmail.com

*Irina A. SEKERINA*

*The City University of New York (CUNY), Estados Unidos de América*

irina.sekerina@csi.cuny.edu

DOI: <https://doi.org/>

10.25189/2675-4916.2022.v3.n1.id629

Recepción: 07 Febrero 2022

Aprobación: 16 Junio 2022

Publicación: 26 Septiembre 2022



Acceso abierto diamante

## Abstract

Heritage speakers represent a special category of bilinguals who are exposed to their first language at home in the childhood, but later acquire the main language of their society which becomes dominant. Brazil has numerous communities of heritage speakers of many languages such as Japanese, German, Italian, Polish, and Ukrainian; however, only few speech corpora are being collected. Here we describe the protocol of the data collection and discuss some points about data management for the BraPoRus (Brazilian Portuguese-Russian) corpus, a spoken corpus of heritage Russian in Brazil. The participants are 26 elderly speakers who were born in Brazil or came to Brazil as children in the 1950s. The protocol of the data collection includes: 1) a brief sociodemographic questionnaire; 2) a working memory test in Russian and Brazilian Portuguese using the Month-Ordering task; 3) a semi-spontaneous narrative about the history of the participants' family and their immigration to Brazil; 4) the Bilingual Language Profile; 5) a sociolinguistic interview with 139 questions; 6) unscripted dialogues between heritage speakers in Russian; 7) an intonation task; and 8) a reading task. The BraPoRus corpus contains 167.5 hours of speech recordings and represents a unique collection of heritage Russian spoken in Brazil. We expect that the protocol described in this

work will be useful both for Brazilian linguists who study other heritage languages, and for researchers of heritage Russian in other countries.

**Keywords:** Bilingualism, Spoken Corpus, Russophones, Heritage Russian, Brazilian Portuguese.

## Resumo

Os falantes de herança representam uma categoria especial de bilíngues que adquirem a sua primeira língua no ambiente familiar na infância, e a sua segunda língua, que com o tempo se torna dominante, na sua vida em sociedade. No Brasil, há muitas comunidades que falam línguas de herança como japonês, alemão, italiano, polonês e ucraniano. No entanto, poucos trabalhos de preservação dessas línguas, com a coleta de corpus, estão sendo conduzidos. No presente trabalho, descrevemos o protocolo de coleta do BraPoRus, o corpus de fala em russo como língua de herança no Brasil. Os participantes são 26 falantes idosos nascidos no Brasil ou que vieram ao país nos anos 1950 ainda crianças. O protocolo de coleta de dados inclui: 1) um curto questionário sociodemográfico; 2) teste de memória de trabalho em russo e em português; 3) uma narrativa semi-espontânea sobre a história da família do participante e a sua imigração ao Brasil; 4) questionário de caracterização de perfil bilíngue; 5) entrevista sociolinguística com 139 perguntas; 6) diálogos em russo entre os falantes de herança sobre temas livres; 7) tarefas de leitura de enunciados com variados padrões entoacionais; 8) tarefa de leitura de texto. O corpus BraPoRus contém 167.5 horas de gravações de fala e representa um acervo único de russo como língua de herança no Brasil. Esperamos que o protocolo descrito neste trabalho seja útil tanto para linguistas brasileiros que estudam outras línguas de herança, quanto para pesquisadores que estudam russo como língua de herança em outros países.

**Palavras-chave:** Bilinguismo, Corpus Oral, Russófonos, Russo como Língua de Herança, Português Brasileiro.

## Introduction

Heritage speakers represent a special category of bilinguals who are exposed to their first language at home in the childhood, but later acquire the main language of their society that becomes dominant (CUMMINS, 2005; POLINSKY; KAGAN, 2007). Heritage languages are acquired under conditions of variable qualitative and quantitative input, and the linguistic competences of the heritage speakers change over the course of their lifespan (D’ALESSANDRO; NATVIG; PUTNAM, 2021; MONTRUL, 2016).

Brazil has numerous communities of heritage speakers of many languages such as Japanese (ABREU MORATO, 2011); German, including a rare dialect Hunsrückisch (GEWEHR-BORELLA; ZIMMER; ALVES, 2011); Italian, including a north Italian dialect Talian/Venetian (BENINCÁ, 2018); Polish (MILESKI, 2017), and Ukrainian (COSTA; LOREGIAN-PENKAL, 2015). However, in Brazil, only few speech corpora are being collected in heritage language studies. The best-known Brazilian corpus is the ALMA-H (Atlas Lingüístico-Contatual das Minorias Alemãs na Bacia do Prata: Hunsrückisch, Linguistic Contact Atlas of German Minorities in the Rio de La Plata Basin: Hunsrückisch) for Hunsrückisch (ALMA-H, 2022; ALTENHOFEN; MORELLO, 2018). However, the data are not open, the methodology section on the site mentions a number of questionnaires not available for download. The grammar and readings tasks are especially developed for Hunsrückisch and standard German.

Here we describe the protocol of data collection and discuss some points about data management for the BraPoRus (Brazilian Portuguese-Russian) corpus, a spoken corpus of heritage Russian in Brazil. The participants are 26 elderly speakers who were born in Brazil or came to Brazil in the 1950s as children. This corpus is being recorded since February, 2021, when the pandemic COVID-19 restrictions made it mandatory to switch to online data collection for linguistic research projects. The protocol of the data collection contains a variety of tasks developed by researchers from different areas, from phonetics to psycholinguistics. The tasks include: 1) a brief demographic questionnaire; 2) a working memory test in Russian and Brazilian Portuguese using the Month-Ordering task (KEMPLER et al., 1998; GORAL et al., 2011); 3) a semi-spontaneous narrative about the history of the participants’ family and their immigration to Brazil; 4) the Bilingual Language Profile (BLP) (BIRDSONG; GERTKEN; AMENGUAL, 2022); 5) a sociolinguistic interview with 139 questions adapted from the long HLVC (Heritage Language Variation and Change, Toronto) questionnaire (NAGY, 2016); 6) unscripted dialogues between heritage speakers in Russian; 7) an intonation task; and 8) a reading task. In our recent work, we confirmed that the acoustic quality of the BraPoRus remote recordings is satisfactory making them adequate for basic acoustic analysis and automatic transcription (SEKERINA et al., in revision to resubmit).

## 1. THEORETICAL BACKGROUND

### 1.1. HERITAGE RUSSIAN SPEAKERS IN BRAZIL

The history of Russophone immigration to Brazil is described in a number of dissertations, theses and books (BYTSENKO, 2006; CHNEE, 2016; HIGA, 2015; RUSEISHVILI, 2016; VOROBIEFF, 2006; ZABOLOTSKY, 1998). The first immigrants from the Russian Empire arrived in Brazil at the turn of the 19th century to the 20th; most of them were Russian German and religious minorities immigrants, such as Old Believers (BYTSENKO, 2006). After the Bolshevik Revolution, four Russian immigration waves took place in Brazil comprising the following categories of the immigrants: 1) 1921 – 1941: the civil and military refugees from the former Russian Empire; 2) the post Second World War period: Soviet Displaced Persons (DPs) and the reminiscent families of the “white” Russian community in Europe; 3) the 1950s: Russian “white” stateless refugees from China; 4) since 1991: the contemporary immigration initially stimulated mainly by economic reasons (RUSEISHVILI, 2016; SMIRNOVA HENRIQUES; RUSEISHVILI, 2019; SKOROBOGATOVA et al., 2021; VOROBYEVA; ALESHKOVSKI; GREBENYUK, 2018).

We presented the relevant IBGE statistics about the number of Russians who had entered Brazil until the 1950s in our recent report (SKOROBOGATOVA et al., 2021). The Russian immigrants that arrived in Brazil from China, mainly Harbin, in the 1950s, present a special interest for our study because the Russian community in China until the 1950s was very strong, and this created the unique conditions for the Russian language preservation there (OGLEZNEVA, 2009). In addition, nearly half of the Chinese Russians that arrived in Brazil were infants and young people under 29 years (RUSEISHVILI, 2018). In São Paulo, Russian schools for these children functioned until the 1970s (HIGA, 2015; SKOROBOGATOVA et al., 2021; VOROBIEFF, 2006). We estimate that up to 1.500 of these Chinese Russians could still be alive (SKOROBOGATOVA et al., 2021) and, at least to some degree, continue to speak Russian.

## 1.2. BILINGUALISM, AGING, AND MORIBUND HERITAGE LANGUAGES

Language dominance is dynamic and changes across the life course in all the bilinguals (GROSJEAN; LI, 2013). The heritage speakers pass through the stages of dominance in their heritage language in the childhood, more or less balanced bilingualism at the intermediate stage, and finally enter the stage when the proficiency in the heritage language is dwindling (AALBERSE; BACKUS; MUYSKEN, 2019). When the population of heritage speakers ages, most of them find themselves at this last stage when they use their heritage language mainly for home and family topics.

In addition, the speech of elderly people can bring some age-specific characteristics that do not come from their bilingual conditions even when the aging is healthy (BOT; MAKONI, 2005). To characterize the age effects on the speech and to study the language variation and communication abilities in later life, corpora of speech of elderly people have recently started to be collected (BOLLY; BOUTET, 2018). The aging lexicon is affected by both environmental exposure and several cognitive mechanisms associated with learning, representation, and retrieval of information (WULLF et al., 2019). Aging in multilingual contexts requires additional factors to be considered (BOT; MAKONI, 2005), and represents a new exciting area to which the BraPoRus corpus can make an important contribution.

A special category of the moribund (endangered) heritage languages contains the languages that are spoken by elderly individuals who represent a last generation of proficient speakers: in this case, the language dies together with these last speakers (D'ALESSANDRO; NATVIG; PUTNAM, 2021). While Russian is not a moribund language, the descendants of the Russian emigrants that left Russia after the Bolshevik revolution speak its frozen variety that preserves some features of the language from one hundred years ago. Oglezneva (2009) shows that Russian spoken in Harbin in the first half of the 20th century preserves some archaic words, and says that the term “archaic” is relative: while some words are considered archaic in Russia, they are currently used abroad by the Russian emigrants and can be important for the self-identification in the community. She gives examples of using the old Russian units of measurements as “verst”, “sazhen”, “pood” and “funt” instead of the metric system terms established in Russia as mandatory in 1917.

D'Alessandro et al. (2021) highlight three groups of challenges faced by researchers who work with moribund heritage languages. First, the baseline challenge: a baseline standard spoken by previous generations cannot be established. When a baseline is not established, it can be hard to separate the heritage features of the spoken language and the influence of the society language (AALBERSE; BACKUS; MUYSKEN, 2019). Any heritage speakers should be studied in the context of their bilingual or multilingual conditions. Most heritage Russian speakers in Brazil speak Portuguese most of the time, this means that their Russian can suffer some Portuguese influence at the pronunciation level and even to get some grammar features typical for Russian as L2. These features can be especially notable in the speakers who were born in Brazil and did not go to a Russian school. However, the relevant baseline for BraPoRus can be set up by using the recordings of the speech of elderly Russians in Harbin performed by Oglezneva (2009) in 2000-2002. The Russian language spoken in Harbin has some special features and exhibits some influence from Chinese; nonetheless, comparisons of the recordings made in Harbin in the

21st century and in the community of Chinese Russians in Brazil who left China in the 1950s can provide a useful reference.

The second challenge identified by D'Alessandro et al. (2021) is data elicitation: usually, the number of speakers is low, they are elderly and unfamiliar with the actual data collection methods. In the case of BraPoRus, the pandemic COVID-19 restrictions favored online data collection: in February, 2021, the elderly people in Brazil were in lockdown at home, had a lot of time, and many of them learned how to use Zoom for various purposes, from speaking to their family to participation in Orthodox church meetings.

We consider that the difficulties with the definition of the language dominance and proficiency in a heritage language also suits the data elicitation challenge. Many heritage speakers are not able to read, so they cannot perform the tasks based on reading common for the proficiency level definition (AALBERSE; BACKUS; MUYSKEN, 2019). Although there is a big variety of proficiency tests for English, not many options exist for Russian. So, we decided to follow the Bilingual Language Profile protocol based on the self-evaluation and not related to reading (BIRDSONG; GERTKEN; AMENGUAL, 2022).

The third challenge identified by D'Alessandro et al. (2021) is the data amount challenge, and it coalesces from the first two: a limited number of data does not provide a base to obtain a good statistical power. We have currently collected for BraPoRus 167.5 hours of speech recordings from 26 participants, and the data collection is going on (SMIRNOVA HENRIQUES et al., 2021; 2022).

### 1.3. COLLECTING HERITAGE LANGUAGE DATA

Data collection practices for corpus construction include collecting spontaneous speech, collecting questionnaires, and data elicitation (AALBERSE; BACKUS; MUYSKEN, 2019). The challenges of data collection begin with the Observer's Paradox: the knowledge of being recorded alters the way of speaking (LABOV, 1972). However, in the situation of the online/phone interaction, the recording process can be perceived by participants as less intruding, because they stay in their homes and not in artificial laboratory conditions.

In the situation of the heritage language studying, there is an additional pressure coming from the interviewer because the researchers generally speak the contemporary variety of the language. In the ideal situation, the interviewer should be a heritage speaker; however, it is not feasible when dealing with elderly people, and their language variety is moribund. To create more natural conditions for the conversation and to study everyday language use, it is important to organize some spontaneous dialogue interactions between the heritage speakers (AALBERSE; BACKUS; MUYSKEN, 2019; JOHANNESSEN, 2021).

The common issues in data curation (storage, transcription and annotation) also can look quite differently in the context of heritage language studies. First, when the conversations are about the family history, the data can be framed in two ways: as an oral corpus which are generally anonymized, or as a database of oral history interviews (SMIRNOVA HENRIQUES et al., 2021). The question of how to balance the participant anonymization and the preservation of the memories is very important. For example, the Code of Ethics of TalkBank, a popular database for open sharing of speech corpora, establishes that only age and location of the oral corpus recordings could be publicly available (MACWHINNEY, 2022). Many participants of the BraPoRus corpus have clearly expressed their desire to preserve their names and history of their parents, and these data have a special importance for the historical and sociological studies of Russian immigration. Columbia University Libraries (2022) made available a number of oral history interviews recorded in the 1960s by Russian immigrants who are identified and talk about their biography mentioning their personal data. However, the recordings made in this format cannot be submitted to the TalkBank database. As far as we know, there are no publicly available annotated heritage speech corpora that contain personal data as the full names and family history, so we need to develop a strategy that permits to share the data, but would not be considered by the Ethics Committee as an attempt to violate the privacy and confidentiality of the research participants who authorized this sharing without foreseeing the consequences. To achieve this goal, we are currently revising

the questionnaire of our sociolinguistic interview with the PUC-SP juridical assistance: probably, a part of the recordings should be kept only for our use.

Another challenge is the transcription procedure. There are a number of tools that perform the automatic transcription in Russian (SEKERINA et al., in revision to resubmit; SMIRNOVA HENRIQUES et al., 2022), but they lose uncommon words and morphological markers especially interesting in the heritage language studies, and do not allow to study the code-switching.

Now we present the protocol used for the BraPoRus data collection.

## 2. METHODS AND RESULTS

### 2.1. PARTICIPANTS

The essential criteria for participant selection were: 1) age 59 years and older (range: 59 - 98); 2) living in Brazil for the most part of their life or being born in Brazil, and speaking Brazilian Portuguese in a native-like way; 3) high proficiency in Russian as a heritage language, sufficient to maintain a conversation for an hour; 4) no long-term residence in Russia; 5) no documented cognitive impairment (SMIRNOVA HENRIQUES et al., 2021). At the present moment, we have recorded 167.5 hours of speech collected from 26 participants with a mean age of 75.7 years.

Data collection for the postdoctoral project of Dr. Smirnova Henriques “Russophones and Brazilian Portuguese: an interdisciplinary study” was approved by the Ethics Committee of Pontificia Universidade Católica de São Paulo (CAAE 09079219.9.0000.5482). This approval authorized the recording of participants’ speech for analyses in our laboratory. However, most participants desired to make turn their recordings publicly available to share the history of their families. We are currently asking for a reconsideration of the project to apply a new form of the informed consent and to obtain authorization of the Ethics Committee for the public use of these data, as desired by the participants.

### 2.2. RECORDING PROCEDURE

Under the pandemic COVID-19 restrictions, the participants were recorded through the video teleconferencing software Zoom<sup>TM</sup> (16 participants) or through a phone call (10 participants). The recording sessions were conducted according to the availability of each participant, and lasted between 30 minutes and two hours. The maximal number of sessions until now was ten. The Zoom conferences were recorded on the Dell Inspiron PC, with the audio being recorded in .mp3 format, and the video, in .mp4. The phone calls were recorded on the Xiaomi Redmi 6 smartphone, using the Android package program, also in .mp3 format. As of now, a total of 167.5 hours of the speech have been recorded (Table 1).

**Table 1**

Duration in minutes and number of participants for each recorded task of the BraPoRus corpus Extracted from Table 2 in the manuscript by SEKERINA et al in revision to resubmit

Task	Monologue	BLP	Sociolinguistic interview	Intonational phrases	Reading text	Dialogue
Number of participants	26	21	21	14	10	5
Duration (min)	2,011	1,910	5,636	116	107	272

SEKERINA et al (in revision to resubmit)

### 2.3. TASKS FOR THE BRAPORUS DATA COLLECTION

In this section, we detail the tasks used for the BraPoRus data collection.

**Sociodemographic questionnaire.** A brief sociodemographic questionnaire was administered in Russian during the first contact by phone call without recording; the interviewer wrote down the name of the participant and their parents, age, date and place of birth, age of arrival in Brazil, profession, familiarity with Zoom, and contacts with other heritage Russian speakers. During this conversation, the interviewer also informally assessed the ability to keep the conversation in Russian.

**Working memory test.** For verbal working memory assessment, we applied the Months-Ordering task (KEMPLER et al., 1998; GORAL et al., 2011) in Russian and Portuguese. The participants listen to an increasingly long set (from 2 to 7) of months, presented out of calendar order, and are asked to recall the months back in the order they appear in the calendar. This task was chosen to avoid the influence of the vocabulary size in unbalanced bilinguals. This protocol is described in detail in our recent report (SKOROBOGATOVA et al., 2021). In the group of elderly heritage Russian-Brazilian Portuguese bilinguals, the median working memory score in Russian (3 months) was 1.5-fold lower than in Portuguese (4.5 months) (SKOROBOGATOVA et al., 2021).

**A semi-spontaneous narrative about the history of the participants' family and their immigration to Brazil.** We recorded a semi-spontaneous narrative about the history of the participants' family and their immigration to Brazil (monologue) for 26 participants (Table 1). The interviewer spoke the contemporary Russian as L1 and Brazilian Portuguese as L2. The participants were informed that they are recorded and asked to speak only Russian. The interviewer allowed the speaker to tell whatever he/she remembers and only resorted to additional questions when the participant stopped talking. This task lasted for 1-3 sessions.

**Bilingual Language Profile (BLP).** We evaluated the language dominance of elderly heritage Russian-Brazilian Portuguese bilinguals using the Bilingual Language Profile questionnaire in Russian (BIRDSONG; GERTKEN; AMENGUAL, 2022). It consists of four parts: language history, use, proficiency, and attitudes. Each part relies on self-evaluation and includes the questions with Likert scale answers. We administered the questionnaire orally via a previously scheduled phone call or Zoom videoconference, and audio recorded the answers to incorporate them in the BraPoRus corpus as speech samples (SKOROBOGATOVA; SMIRNOVA HENRIQUES; MADUREIRA, 2021). We strictly followed the BLP protocol, and calculated the global language dominance score for each language as a sum of four scores for language history, use, proficiency, and attitudes. The final language dominance score was obtained through the subtraction of the Russian global language score from the Portuguese global language score. The final score can range from -218 (dominant Russian) to 218 (dominant Portuguese).

The preliminary data on the language dominance score evaluation in 23 BraPoRus participants are shown in Figure 1 (SKOROBOGATOVA; SMIRNOVA HENRIQUES; MADUREIRA, 2021). Most of heritage Russian speakers, 18 out of 23, were dominant in their society language, Portuguese (Figure 1A). In this group, seven participants were born in Brazil, other eleven arrived at the mean age of 7.8 years. The data obtained from other five participants are shown on Figure 1B. One of them was a balanced bilingual (the final dominance score -1.45, age of arrival 10 years). Four participants were dominant in Russian, their ages of arrival range from 14 to 17 years. A strong negative correlation was found between the age of arrival to Brazil and BLP score (Spearman  $R = -0.76$ ,  $p = 0.0006$ ).

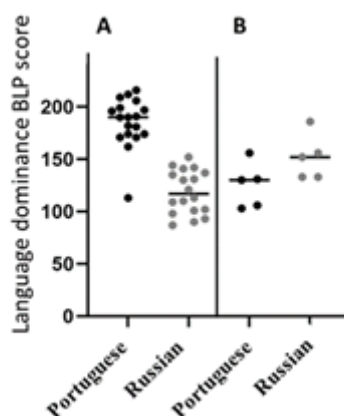


Figure 1

Language dominance profiles in elderly heritage Russian-Brazilian Portuguese bilinguals with dominant Portuguese (A) and dominant Russian (B). The x-axis shows the language evaluated for global dominance score. Horizontal lines show the median scores for each language in each group. Reproduced from the 29th SIICUSP proceedings (SKOROBOGATOVA; SMIRNOVA HENRIQUES; MADUREIRA, 2021)

**A sociolinguistic interview.** A sociolinguistic interview contained 139 questions selected from the long HLVC (Heritage Language Variation and Change, Toronto) project questionnaire (NAGY, 2016) and translated from English into Russian. The original HLVC protocol consists of 366 questions separated into the following sections: Demography, Neighbourhood, Socializing, Events, Childhood, School, Work, Dating/Marriage, Dreams, Fear, Danger, Gardening, Driving, Family Traditions, Ethnic groups, and Language. We kept most of the topics and added new questions including the following issues: 1) experience of travelling to Russia (if any), 2) contacts with friends and family in Russia, 3) documents used to enter Brazil, 4) cooking (the Russian cuisine or others), 5) literature and musical preferences, 6) preservation of the Russian language (how it was preserved by the family; if children and grandchildren speak/understand Russian), 7) balance of Russian and Portuguese in the participants' life, 8) difference between Russian spoken in the participants' family and Russian spoken by other immigrants, 9) written materials and pictures preserved by the family, 10) folk tales and songs performed by the parents.

As well as for the monologue task, the interviewer spoke contemporary Russian as L1 and Brazilian Portuguese as L2. The interviewer was instructed to allow the participant to speak and not to hurry with the formulation of the next question. Only few participants have completed this task until now because it takes many sessions.

**Unscripted dialogues between heritage Russian speakers.** We introduced this task to minimize the observer's influence and to record the conversation in more natural conditions. As of now, five heritage Russian speakers took part in dialogues between them (Table 1). The researcher was present at the Zoom conference with muted microphone and closed camera and only recorded the meeting. Before the conversation, the participants received a list of suggested themes given to them in Portuguese to not influence their vocabulary in Russian. The list included such themes as Daily routine, Family, Russian traditions, Hobby, and Life in pandemic COVID-19 conditions. Until now, five dialogues were recorded. The mean duration was 67 minutes.

**Intonation task.** The team of Dr. Skrelin from Saint Petersburg State University constructed the set of 17 trials (phrases or short dialogues) in Russian specifically for research of cross-linguistic influence on intonation between Russian and other languages. The initial set of 11 trials, developed by Nina Volskaya to study the perception of Russian rise-falls by speakers of other languages, was extended to encompass other melodic types that may be specific for the Russian intonation system. The trials were tested on speakers of German (SKRELIN, 2017) and Finnish (unpublished).

Each of the 17 trials contains a target phrase where a particular Russian melodic pattern is expected. In Table 2 below, we show the expected melodic pattern before each target word (the expected main word, or nucleus) in Russian. The melodic patterns are described in terms of the Bryzgunova intonation system of

seven “intonation constructions” (ICs) (BRYZGUNOVA, 1977), as well as in terms of the ToBI-like description suggested by Odé (ToRI) (2008). The trials were translated into Portuguese, and the participants recorded the task both in Russian and Portuguese. Only 14 BraPoRus participants from 26 who were able to read performed the task. The English translation is provided only for clarity in the current article.

**Table 2**

The trials for intonation task in Russian and Portuguese Original text in Russian translation into Portuguese and English The target words in Russian are underlined and preceded by the expected melodic pattern annotation

Trial Number	Russian	Translation to Portuguese	Translation to English
1	Если у вас болит [IC3, H*L]голова, примите аспирин.	Se você está com dor de cabeça, tome uma aspirina.	If you have a headache, take aspirin.
2	Если аспирин не [IC3, H*L]помогает, обратитесь к врачу.	Se a aspirina não ajudar, procure um médico.	If the aspirin doesn't help, see a doctor.
3	Если у вас высокая [IC3, H*L]температура, оставайтесь дома, примите аспирин и вызовите врача.	Se você está com febre, fique em casa, tome uma aspirina e chame um médico.	If you have fever, stay home, take aspirin and call a doctor.
4	Если вы почувствовали запах [IC3, H*L]газа, немедленно сообщите об этом по телефону 112.	Se você sentir cheiro de gás, ligue imediatamente para o número 112.	If you smell gas, immediately call 112.
5	Если пойдёшь в [IC3, H*L]магазин, купи молоко, сыр, помидоры и лук.	Se você for no supermercado, compra leite, queijo, tomates e cebola.	If you go to the supermarket, buy some milk, cheese, tomatoes, and onions.
6	Так как погода испортилась и началась [IC3, H*L]гроза, мы на озеро не поехали.	Como o tempo piorou e começou uma tempestade, acabamos não indo para o lago.	Since the weather turned nasty and a storm started, we didn't drive to the lake.
			When leaving the

10	- Назовите времена года. - Зима, весна, лето. - А [IC4, L*H]осень?	- Me fale as estações do ano. - Inverno, primavera, verão. - E o outono?	- Name the seasons of the year. - Winter, spring, summer. - And autumn?
11	- Хорошо, что скоро отпуск! [IC6, H*N]Море, [IC6, H*N]солнце, [IC6, H*N]пальмы, хорошая [IC6, H*N]компания... Мечта!	- Que bom que as férias estão chegando! Mar, sol, palmeiras, boa companhia... um sonho!	- That's good that vacation is coming soon! Sea, sun, palm trees, nice company... Just like a dream!
12	- Вы [IC3, H*L]видели Петину новую квартиру? - Да. Она маленькая, но очень удобная.	- Você viu o apartamento novo do Pedro? - Sim. É pequeno, mas bem confortável.	- Have you seen Peter's new apartment? - Yes. It is small, but very comfortable.
13	[IC2, HL*]Осторожно! Внизу ступенька!	- Cuidado! Tem um degrau embaixo!	Watch your step as you go downstairs!
14	- Говорят, он стройный [IC3, H*L]брюнет? - Конечно, он стройный. Но брюнетом я бы его не назвал.	- Dizem que ele é magro e tem cabelo escuro? - Claro, ele é magro. Mas não chamaria o cabelo dele de escuro.	- Everybody says he's a slim brunet. - For sure, he's slim. But I wouldn't call him a brunet.
15	- Аня с Мишей живут теперь на Шри-Ланке. - Где-[IC6, H*N]где они теперь живут?	- A Ana e o Michel moram agora no Sri Lanka. - Onde é que eles moram?	- Anna and Misha now live in Sri Lanka. - Where did you say they live?
16	- Какие [H*N]замечательные у Вас [L*]серьги!	- Que brincos lindos você tem!	- What wonderful earrings you have!

Authors

Using this task, we have recently analyzed the intonational interference of Brazilian Portuguese L2, the society language, on heritage Russian L1 in contrastive (elliptical) questions, yes/no questions and echo wh-questions, using the trials 9, 10, 12 and 13 (KACHKOVSKAIA et al., 2022). We observed a strong influence of Brazilian Portuguese on the prosody of heritage Russian, with the number of intonational changes over 70 %, reaching 87 % for yes/no questions and echo wh-questions.

**Reading task.** To study the pronunciation features of the heritage Russian speakers, we asked the BraPoRus participants to read a phonetically representative text of 443 words developed for studying the pronunciation of monolingual Russian speakers (BONDARKO et al., 1993). Ten participants read the text. However, it contained many words that they did not know, and the readers had many difficulties.

## 2.4. ACOUSTIC QUALITY OF THE BRAPORUS RECORDINGS

We have recently analyzed the acoustic quality of 16 BraPoRus samples (8 Zoom<sup>TM</sup> and 8 phone recordings), each one 15 minutes, using a protocol developed by the Group of Studies in Forensic Phonetics (Geff, 2020) for forensic phonetic analysis. Most Zoom<sup>TM</sup> recordings were classified by acoustic quality as AB: F4 was not visible and fricatives not always identifiable, but the major parts of samples had very good quality for acoustic analysis (SEKERINA et al., in revision to resubmit). Most

phone recordings were classified as B or BC, F3 and F4 were not always identifiable, and the sound quality was affected in a more significant manner. Nonetheless, the quality of all this material was sufficient for F0, F1 and F2 measurements making it adequate for basic acoustic analysis.

## 2.5. TRANSCRIPTION AND ANNOTATION OF THE BRAPORUS CORPUS

The BraPoRus has 167.5 hours of recordings, and manual transcription would be very laborious and time-consuming. We have recently tested three automatic transcription tools. Commercially available transcription software Sonix was shown to yield the lowest word error rate (WER) (SMIRNOVA HENRIQUES et al., 2022). Next, we tested automatic transcription by Sonix for four Zoom™ and four phone recorded 15-minutes samples, and found the mean WERs to be 11.51 and 8.21, respectively, considered quite low (SEKERINA et al., in revision to resubmit). However, manual curation is still necessary.

The annotation will include the assignment of word class labels and code-switching tags (BULLOCK et al., 2018; GUZMAN, 2017). The annotation scheme is currently being developed.

## 3. DISCUSSION AND FUTURE DIRECTIONS

In the current article, we present a protocol used for the BraPoRus corpus construction to study heritage Russian speech in Brazil. The protocol describes both the tasks for corpus recording and complementary tasks planned to obtain additional information on language dominance of the participants and the interference of Brazilian Portuguese on heritage Russian speech at the levels of intonation or pronunciation. At this point, BraPoRus contains 167.5 hours of recordings collected from 26 elderly heritage Russian speakers (Table 1). Here we do not describe the participants' demographic characteristics in detail because this information has already been provided elsewhere (SMIRNOVA HENRIQUES et al., 2021), and because in this article we focus on the procedure of corpus construction.

First analyses of the BraPoRus project data addressed the working memory in heritage Russian-Brazilian Portuguese bilinguals (SKOROBOGATOVA et al., 2021), language dominance profiles (SKOROBOGATOVA; SMIRNOVA HENRIQUES; MADUREIRA, 2021) and sociodemographic characterization (SMIRNOVA HENRIQUES et al., 2021) of these speakers. The intonation interference of L2 Brazilian Portuguese on L1 heritage Russian also was described (KACHKOVSKAIA et al., 2022). The analyses of the BraPoRus corpus, until the current moment, include the evaluation of the acoustic quality of Zoom™ and phone recordings, and WER definition for automatic transcription by Sonix (SEKERINA et al., in revision to resubmit; SMIRNOVA HENRIQUES et al., 2022). Reasonable acoustic quality and low WER for automatic transcription show the feasibility of data collection for corpus building online even for elderly heritage speakers. However, the manual curation after automatic transcription is necessary, and the data are not appropriate for acoustic analyses that involve the F3 and F4 evaluation.

The main challenge during the BraPoRus data collection online consists in the organization of dialogues between heritage Russian speakers to record their interactions without the observer's influence, as performed for the CANS corpus (JOHANNESSEN, 2021). The BraPoRus participants are elderly and not very interested in making friends on the internet. The best interaction was observed between two friends who knew one another for more than 60 years. Another challenge is performing tasks that require reading skills, such as reading intonation tasks and phonetically representative text: only 14 out of 26 participants were able to carry out these activities. The recordings of more speakers and new sessions continue at the same time as the already recorded speech is automatically transcribed and manually verified. The next steps include annotation and decision in which format the data will be made available.

The most serious challenge for the corpus construction as a whole is to obtain the authorization of the Ethics Committee to publicly share the corpus recordings: they were initially collected within a broader project of studying Russian – Brazilian Portuguese bilingualism that did not suppose the open sharing (SMIRNOVA HENRIQUES et al., 2020). Although many participants wish to reveal their family history,

the respect of the privacy and confidentiality is the mandatory norm requested by the Brazilian and foreign Ethics Committees (MACWHINNEY, 2022; PUC-SP, 2022). While in History and Sociology the names can be part of the open data (Columbia University Libraries, 2022), this is not commonly applied to the voice samples in linguistic research.

Our main reference on the initial step of the research was the project of Heritage Language Variation and Change in Toronto (NAGY, 2016) that suggested extensive questionnaires for interviewing the heritage speakers. However, the site of the corpus has not functioned during the last months or, at least, is no longer accessible from Brazil. Another important reference is the CANS (Corpus of American Nordic Speech) that prioritized to diminish the interviewer's influence (JOHANNESSEN, 2021). The BraPoRus corpus contains many additional tasks suggested by researchers in our group with different backgrounds, and this allows us to characterize the population of heritage Russian-Brazilian Portuguese bilinguals much better. As far as we know, there is no other oral corpus of elderly heritage Russian speakers.

For the next steps, we also plan to assess narrative abilities of the elderly heritage Russian speakers in Russian and Brazilian Portuguese using Multilingual Assessment Instrument for Narratives (MAIN) (GAGARINA et al., 2016). This is a widely used tool available for testing the narrative competence in more than 80 languages, in children and adults. MAIN was specifically created to assess development of narrative skills in bilinguals, and is based on multidimensional model of story organization and six-picture stories controlled for cognitive and linguistic complexity. We are currently adapting the materials for remote testing in Russian and Brazilian Portuguese (SKOROBOGATOVA; SMIRNOVA HENRIQUES; GAGARINA, 2021). The face-to-face protocol has been translated to Portuguese previously (CUNHA DE AGUIAR; MARTINS DOS REIS, 2020), but it has never been applied in Brazil.

The BraPoRus corpus provides a huge quantity of data on the heritage Russian in Brazil spoken by elderly people, and heritage Russian-Brazilian Portuguese bilingualism. We expect that the protocol described in this work will be useful both for Brazilian linguists that study other heritage languages, and for researchers of heritage Russian in other countries.

## ADDITIONAL INFORMATION

### CONFLICT OF INTEREST

The authors have no conflicts of interest to declare.

### STATEMENT OF DATA AVAILABILITY

**Anna Smirnova Edelbuettel declares on behalf of all authors of the manuscript “BRAPORUS, SPOKEN CORPUS OF HERITAGE RUSSIAN IN BRAZIL: PROTOCOL OF DATA COLLECTION”, submitted to *Cadernos de Linguística*, that the aforementioned manuscript is submitted as a project registration and, therefore, does not yet have the analyzed speech data that can be accessed by other researchers. The BraPoRus corpus, described in the article, contains 167.5 hours of recordings, however, at the moment, the authorization of the Ethics Committee of PUC-SP that we have does not allow us to make the recordings publicly available because they contain the personal information of the participants. The data collection protocols used for the recordings are either cited in the manuscript, or made available in the text itself, as explained below:**

1. sociodemographic questionnaire; a semi-spontaneous narrative about the history of the participants' family and their immigration to Brazil; a sociolinguistic interview; unscripted dialogues between heritage Russian speakers; intonation task – described in the manuscript;
2. working memory test - SKOROBOGATOVA, Aleksandra S.; SMIRNOVA HENRIQUES, Anna; RUSEISHVILI, Svetlana; SEKERINA, Irina; MADUREIRA, Sandra. Verbal working memory assessment in Russian-Brazilian Portuguese bilinguals. *Cadernos de Linguística*, v. 2, n. 4, p. 01-24, e572, 2021. DOI 10.25189/2675-4916.2021.V2.N4.ID572.
3. Bilingual Language Profile - <https://sites.la.utexas.edu/bilingual/>
4. reading task - BONDARKO, Liya V. et al. (Orgs.). Prilozhenie №3 k Bulletinu Foneticheskogo fonda russkogo yazika “Fond zvukovykh edinit russkoi rechi” [Attachment №3 to the Bulletin of the Phonetic Bank of the Russian language “Bank of the Russian Speech Sound Units”]. St Petersburg/ Bochum, 1993.

## FUNDING SOURCES

Dr. Smirnova Henriques is supported by postdoctoral fellowship PNPD/CAPES (Programa Nacional de Pós-Doutorado da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior, the process number 88882.315378/2019-01). Aleksandra S. Skorobogatova is supported by the undergraduate fellowship program from FAPESP (Fundação de Amparo à Pesquisa do Estado de São Paulo, the process number 2022/01119-0)

## REVIEW

REVIEWER 1: Ignacio Miguel Palacios Martinez

ORCID: <https://orcid.org/0000-0001-9202-9190>

AFFILIATION: Universidade de Santiago de Compostela, Galiza, Spain.

REVIEWER 2: Marcia dos Santos Machado Vieira

ORCID: <https://orcid.org/0000-0002-2320-5055>

AFFILIATION: Universidade Federal do Rio de Janeiro, Rio de Janeiro, Brazil.

ROUND 1

REVIEWER 1

2022-02-22 | 11:13

This is an interesting paper on the instruments and procedures followed for the compilation of a spoken corpus of heritage Russian in Brazil. In this respect, the paper fulfills the purpose pursued although one would expect to find in the same article some references to the transcription conventions followed together with some information about the annotation scheme. Some mention should also be made to the problems found in the collection of the data. The authors should also make it clear up to what extent this corpus is different from other heritage corpora of the kind. There should also be in the final section an indication of what sort of projects could be carried out with this material.

The paper is written in good academic English and it complies with the conventions typical of this genre. There is a good use of bibliographical references which are complete and well cited.

Some further suggestions follow:

p. 3. In the current work, delete comma

p. 4. arrived to Brazil> arrived in Brazil

p. 4. XIX th century > nineteenth century; XXth century> twentieth century

p. 4. the data collection continues> the data collection still continues

p. 4. he generally speaks> they generally speak

p. 5. there is a number> there are a number

p. 6. In this task, > delete comma

p. 8. Please, revise the caption for the figures. Some of the current information should be included in the body of the text. This also applies to Table 2, p. 10

p. 9. the observer influence> the observer's influence

References

Cunha de Aguiar. Please, alphabetise.

REVIEWER 2

2022-05-07 | 11:32

First of all, many thanks to the authors for this manuscript! Certainly, the protocol about data collection and management described in this project registration will be useful for linguists in Brazil and beyond.

I made few observations in the manuscript and make here some suggestions:

1) the form of listing references (;/&);

2) non plural or plural for references related to more than one author (Org./Orgs.);

3) MADUREIRA, Ssandra;

Rethink/restructure/(re)explain:

4) the idea of "special feature" conceptualized in these words: "The special feature of this corpus is that it is not restricted to the sociodemographic data and speech recordings of the participants, but contains a variety of tasks developed by researchers from different areas, from phonetics to psycholinguistics." > Why is it special? And putting in perspective/profiling which audience (for BraPoRus)?

5) the perception of on-line recording process: "However, in the situation of the online/phone recordings, the recording process is perceived as less intruding even though the informed consent is signed." Perceived by whom? Which evidences are the basis of such perception description?

6) the idea of "never-to-accessing" participants: "If we treat these data as a spoken corpus for linguistic research, the identification of participants would be never permitted" > Is it a practice that is, than, claimed by ethics committees (made of researchers, from different disciplines) or is it a necessary or mandatory condition? Support the alleged anonymization, perhaps in resolutions of ethical conduct in research.

This is a good manuscript of project registration.

ROUND 2

REVIEWER 1

2022-05-30 | 09:52

I believe the comments and suggestions made in my previous review have been successfully incorporated into this revised version so, in my view, it can be now accepted for publication. Some minor points follow for the authors' consideration regarding some aspects of the paper concerning the wording of some sections.

p. 5. "Most Russian heritage speakers in Brazil speak Portuguese the most part of time, so, their Russian can suffer some Portuguese influence"> the most part of the time: this means that

p. 7. "mentioning their personal data, however,"> mentioning their personal data; however,

p. 11. "The interviewer was instructed to allow the participant to speak and to not be hurry to make the next question"> and not to hurry with the formulation of the next question

p. 15: "The main challenges during the BraPoRus data collection online is organization of dialogues between Russian heritage speakers to record" data collection online was the organization of dialogues...

p. 15: " The BraPoRus participants are elderly and not very interested in making new friends by internet"> inrested in making friends on line or on the internet

Please make sure that the bibliographical references are complete. Check the date of publication for each of them.

REVIEWER 2

2022-06-07 | 05:37

This project registration is a good contribution to literature and a useful material for who has interest in database, in data curation and in heritage language, as it reports the protocol of the data collection of one and explores important points about data management for a spoken corpus of heritage Russian in Brazil,

for BraPoRus, as well as ethical points related to care for the image, family history and recording conditions of the participants (elderly speakers).

## AUTHORS' REPLY

ROUND 1

2022-05-24

Dear reviewer 1,

Thank you a lot for your comments, we carefully revised the text following your suggestions.

As requested, we added detailed information about the transcription process and added essential information about the annotation scheme. The discussion of main problems of our data collection also was included, as well as examples of other heritage corpora and data analyses.

Sincerely,

Anna Smirnova Henriques

Dear reviewer 2,

Thank you for your comments, we carefully revised the text following your suggestions. In relation to the comment 4, we did not focus on profiling audience because the data cannot be publicly available yet. The phrase mentioned in comment 5 was reformulated.

In relation to the comment 6, the section about anonymization was rewritten and completed. The orientations for the preparation of research protocols on the site of the PUC-SP Ethics Committee (<https://www.pucsp.br/cometica/orientacoes-para-elaboracao-de-protocolo-de-pesquisa>) directly mentions the requirement to respect the privacy and confidentiality of the research participants. As requested, the informed consent form should ensure the maintenance of confidentiality and privacy of the research subject's data, before, during and after the end of the research. The international data protection regulation for speech recordings is discussed in some recent articles ([https://www.researchgate.net/publication/336957804\\_Preserving\\_privacy\\_in\\_speaker\\_and\\_speech\\_characterisation](https://www.researchgate.net/publication/336957804_Preserving_privacy_in_speaker_and_speech_characterisation)). We are currently discussing with the PUC-SP Ethics Committee the possibility to open the corpus when the participants authorize the open use of data. The Ethics Committee sent our request to the juridical assistance, but we do not have the final decision yet.

Sincerely,

Anna Smirnova

## REFERENCES

- AALBERSE, Suzanne; BACKUS, Ad; MUYSKEN, Pieter. Heritage languages: a language contact approach. Amsterdam/Philadelphia: John Benjamins Publishing Company, 2019. DOI <https://doi.org/10.1075/sibil.58>.
- ABREU MORATO, Geanne Alves de. Situando a língua japonesa no contexto da história do ensino de línguas no Brasil. HELB, v. 5, n. 5, article 4, 2011. Available at <http://www.helb.org.br/index.php/revista-helb/ano-5-no-5-12011/190-situando-a-lingua-japonesa-no-contexto-da-historia-do-ensino-de-linguas-no-brasil>. Accessed on 30 January, 2022.
- ALMA-H. Questionários. Available at <https://www.ufrgs.br/projalma/questionarios/>. Accessed on 29 January, 2022.
- ALTENHOFEN, Cleo Vilson; MORELLO, Rosângela (Orgs.). Hunsrückisch: inventário de uma língua do Brasil. Florianópolis: Garapuvu, 2018. Available at <https://lume.ufrgs.br/handle/10183/194384>. Accessed on 16 November, 2021.
- BENINCÁ, Ludimilla Rupf. Sócio-história do contato entre o vêneto e o português: um estudo de caso. PAPIA: Revista Brasileira de Estudos do Contato Linguístico, v. 28, n. 1, p. 109-132, 2018. Available at <http://revistas.fflch.usp.br/papia/article/view/3029>. Accessed on 30 January, 2022.
- BIRDSONG, David; GERTKEN, Libby M.; AMENGUAL, Mark. Bilingual Language Profile: An Easy-to-Use Instrument to Assess Bilingualism. Available at <https://sites.la.utexas.edu/bilingual/>. Accessed on 29 January, 2022.
- BOLLY, Catherine T.; BOUTET, Dominique. The multimodal CorpAGEst corpus: Keeping an eye on pragmatic competence in later life. Corpora, v. 13, n. 3, p. 279-317, 2018. DOI <https://doi.org/10.3366/cor.2018.0151>.
- BONDARKO, Liya V. et al. (Orgs.). Prilozhenie №3 k Bulletinu Foneticheskogo fonda russkogo yazika "Fond zvukovykh edinit russkoi rechi" [Attachment №3 to the Bulletin of the Phonetic Bank of the Russian language "Bank of the Russian Speech Sound Units"]. St Petersburg/ Bochum, 1993.
- BOT, Kees de; MAKONI, Sinfree. Language and aging in multilingual contexts. Clevedon / Buffalo / Toronto: Multilingual Matters LTD, 2005.
- BRYZGUNOVA, Elena A. Zvuki i intonaciia russkoi rechi [Sounds and intonation of Russian speech]. Moscow: Russkiy yazyk, 1977.
- BULLOCK, Barbara E.; SERIGOS, Jacqueline; TORIBIO, Almeida Jacqueline; WENDORF, Arthur. The challenges and benefits of annotating oral bilingual corpora: The Spanish in Texas Corpus Project. Linguistic Variation, v. 18, n. 1, p. 100-119. DOI <https://doi.org/10.1075/lv.00006.bul>
- BYTSENKO, Anastassia. Imigração da Rússia para o Brasil no início do século XX. Visões do paraíso e do inferno. 134 p. Dissertation (Master's degree in Language Studies) - Faculdade de Filosofia, Letras e Ciências Humanas, Universidade de São Paulo, São Paulo, 2006. Available at <https://www.teses.usp.br/teses/disponiveis/8/8155/tde-12112007-132926/pt-br.php>. Accessed on 13 August 2021.
- CHNEE, Igor. Imigrantes russos no Brasil e seus descendentes. São Paulo: Igor Schnee, 2016.
- Columbia University Libraries, Digital Collections, Radio Liberty. Available at: [https://dlc.library.columbia.edu/catalog?utf8=%E2%9C%93&search\\_field=all\\_text\\_teim&q=Radio%20Liberty](https://dlc.library.columbia.edu/catalog?utf8=%E2%9C%93&search_field=all_text_teim&q=Radio%20Liberty). Accessed on: 07 February 2022.
- COSTA, Luciane Trennephol da; LOREGIAN-PENKAL, Loremi. A coleta de dados do banco VARLINFE - Variação Linguística de Fala Eslava: peculiaridades e características. Revista Conexão da

UEPG, v. 11, n. 1, p. 102-111, 2015. Available at <https://www.revistas2.uepg.br/index.php/conexao/article/view/6874>. Accessed on 16 November, 2021.

- CUMMINS, Jim. A proposal for action: Strategies for recognizing heritage language competence as a learning resource within the mainstream classroom. *The Modern Language Journal*, v. 89, n. 4, p. 585-592, 2005. Available at <https://www.jstor.org/stable/3588628>. Accessed on 19 November, 2021.
- CUNHA DE AGUIAR, Laís Vitória; MARTINS DOS REIS, Micaela Nunes. Adapting MAIN to Brazilian Portuguese. *ZAS Papers in Linguistics*, v. 64, p. 183-87, 2020. DOI <https://doi.org/10.21248/zaspil.64.2020.572>.
- D’ALESSANDRO, Roberta; NATVIG, David; PUTNAM, Michael T. Addressing Challenges in Formal Research on Moribund Heritage Languages: A Path Forward. *Frontiers in Psychology*, v. 12, 2021. DOI <https://doi.org/10.3389/fpsyg.2021.700126>
- GAGARINA, Natalia; KLOP, Daleen; TSIMPLI, Ianthi M.; WALTERS, Joel. Narrative abilities in bilingual children. *Applied Psycholinguistics*, v. 37, n. 1, p. 11-17, 2016. DOI <https://doi.org/10.1017/S0142716415000399>
- GEFF (Grupo de Estudos em Fonética Forense). Protocolo de análise fonético-forense. In: *Análise Fonético-Forense: Em Tarefas de Comparação de Locutor*. Campinas: Millenium Editora, 2020. p. 3-15.
- GEWEHR-BORELLA, Sabrina; ZIMMER, Márcia Cristina; ALVES, Ubiratã Kickhöfel. Transferências grafo-fônico-fonológicas: uma análise de dados de crianças monolíngues (Português) e bilíngues (Hunsrückisch-Português). *Gragoatá*, v. 16, n. 30, p. 201-219, 2011. DOI <https://doi.org/10.22409/gragoata.v16i30.32931>.
- GORAL, Mira; CLARK-COTTON, Manuella; SPIRO, Avron Ill; OBLER, Loraine K.; VERKUILEN, Jay; ALBERT, Martin L. The Contribution of Set Switching and Working Memory to Sentence Processing in Older Adults. *Experimental Aging Research*, v. 37, n. 5, p. 516-538, 2011. DOI <https://doi.org/10.1080/0361073X.2011.619858>.
- GROSJEAN, François; LI, Ping. *The psycholinguistics of bilingualism*. New Jersey: Wiley-Blackwell, 2013.
- GUZMAN, Gualberto; RICARD, Joseph; SERIGOS, Jacqueline; BULLOCK, Barbara E.; TORIBIO, Almeida Jacqueline. Metrics for modeling code-switching across corpora. In: *INTERSPEECH 2017. Annals [...]*. Stockholm, 2017. p. 67 – 71. <http://dx.doi.org/10.21437/Interspeech.2017-1429>.
- HIGA, Bárbara Silva. O instituto São Vladimir e a presença russa em Santos, pela voz dos imigrantes (1958 – 1968). Term paper (Major in History) – Universidade Católica de Santos, Santos, 2015.
- JOHANNESSEN, Janne Bondi. From Fieldwork to Speech Corpus: The American Norwegian Heritage Language and CANS. In: *Workshop on Immigrant Languages in the Americas (WILA 10)*, 10., 2021. Proceedings [...]. Somerville, MA: Cascadilla Proceedings Project, 2021.
- KACHKOVSKAIA, Tatiana V.; SKRELIN, Pavel A.; GUSEVA, Daria; SKOROBOGATOVA, Aleksandra S.; SMIRNOVA HENRIQUES, Anna; MADUREIRA, Sandra. Cross-Linguistic Influence in Interrogative Intonation Patterns: A Case of Russophones in Brazil. Presentation at the II Congresso Brasileiro de Prosódia. Campinas: UNICAMP, 2022.
- KEMPLER, Daniel; ALMOR, Amit; TYLER, Lorraine K.; ANDERSEN, Elaine S.; MACDONALD, Maryellen C. Sentence comprehension deficits in Alzheimer’s disease: a comparison of off-line vs. on-line sentence processing. *Brain and Language*, v. 64, n. 3, p. 297-316, 1998. DOI <https://doi.org/10.1006/brln.1998.1980>.
- LABOV, William. Some principles of linguistic methodology. *Language in Society*, n. 1, p. 97-120, 1972. DOI <https://doi.org/10.1017/S0047404500006576>.

- MACWHINNEY, Brian. The TalkBank system, 2022. Available at <https://www.talkbank.org/>. Accessed on 05 January 2022.
- MILESKI, Ivanete. Variação no português de contato com o polonês no Rio Grande do Sul: vogais médias tônicas e pretônicas. 2017. 321 p. Thesis (PhD in Language Studies) – Faculdade de Letras da Pontifícia Universidade Católica do Rio Grande do Sul, Porto Alegre, 2017. Available at <http://tede2.pucrs.br/tede2/handle/tede/7794>. Accessed on 16 November, 2021.
- MONTRUL, Silvina. *The Acquisition of Heritage Languages*. Cambridge: Cambridge University Press, 2016. DOI <https://doi.org/10.1017/CBO9781139030502>.
- NAGY, Naomi. Heritage languages as new dialects. In: CÔTÉ, M., KNOOIHUIZEN, J.; NERBONNE, J. (Eds.). *The future of dialects*. Berlin: Language Science Press, 2016. DOI <https://doi.org/10.17169/langsci.b81.81>.
- ODÉ, Cecília. Transcription of Russian intonation, ToRI, an interactive research tool and learning module on the internet. In: HOUTZAGERS, Peter; KALSBECK, Janneke; SCHAEKEN, Jos. (Eds.). *Dutch Contributions to the Fourteenth International Congress of Slavists*. Netherlands: Rodopi, 2008, v. 1. p. 431–450.
- OGLEZNEVA, Elena A. Russkij jazyk v vostochnom zarubezh'e (na materiale russkoj rechi v Harbine) [The Russian language beyond the Eastern frontiers (based on the material in Russian collected in Harbin)]. Blagoveshchensk: Universidade Estatal de Amur, 2009.
- POLINSKY, Maria; KAGAN, Olga. Heritage languages: In the 'Wild' and in the Classroom. *Language and Linguistics Compass*, v. 1, n. 5, p. 368-395, 2007. DOI <https://doi.org/10.1111/j.1749-818X.2007.00022.x>.
- PUC-SP. Orientações para Elaboração de Protocolo de Pesquisa Do Comitê de Ética em Pesquisa da PUC-SP – Sede Campus Monte Alegre. Available at <https://www.pucsp.br/cometica/orientacoes-para-elaboracao-de-protocolo-de-pesquisa>. Accessed on 22 Mat 2022.
- RUSEISHVILI, Svetlana. Ser russo em São Paulo. Os imigrantes russos e a reformulação de identidade após a Revolução Bolchevique de 1917. 2016. 383 p. Thesis (PhD in Social Sciences) - Faculdade de Filosofia, Letras e Ciências Humanas, Universidade de São Paulo, São Paulo, 2016. Available at <https://teses.usp.br/teses/disponiveis/8/8132/tde-13022017-124015/pt-br.php>. Accessed on 16 November, 2021.
- RUSEISHVILI, Svetlana. Perfil sociodemográfico e distribuição territorial dos russos em São Paulo: deslocados de guerra da Europa e refugiados da China após a Segunda Guerra Mundial. *Revista Brasileira de Estudos de População*, v. 35, n. 2, 2018, p. 1-20. DOI <https://doi.org/10.20947/S0102-3098a0036>
- SEKERINA, Irina A.; SMIRNOVA HENRIQUES, Anna; SKOROBGATOVA, Aleksandra S.; TYULINA, Natalia; KACHKOVSKAIA, Tatiana V.; SKRELIN, Pavel A.; RUSEISHVILI, Svetlana; MADUREIRA, Sandra. Brazilian Portuguese-Russian (BraPoRus) Corpus: Transcription and Acoustic Analysis of Elderly Speech During Covid-19 Pandemic. *Linguistic Vanguard*, in revision to resubmit.
- SKOROBGATOVA, Aleksandra S.; SMIRNOVA HENRIQUES, Anna; GAGARINA, Natalia. A elaboração da versão on-line do protocolo MAIN para a avaliação de narrativas de bilíngues russo-português brasileiro. In: CONGRESSO INTERNACIONAL DE PORTUGUÊS COMO LÍNGUA NÃO MATERNA, 3., 2021. *Annals [...]*. Araraquara: Universidade Estadual Paulista, 2021. p. 27.
- SKOROBGATOVA, Aleksandra S.; SMIRNOVA HENRIQUES, Anna; MADUREIRA, Sandra. Bilingual language profiles of the heritage Russian speakers in Brazil, participants of the BraPoRus corpus. In: SIMPÓSIO INTERNACIONAL DE INICIAÇÃO CIENTÍFICA (SIICUSP), 29., 2021. *Annals [...]*. São Paulo: Universidade de São Paulo, 2021. p. 817-820.

- SKOROBOGATOVA, Aleksandra S.; SMIRNOVA HENRIQUES, Anna; RUSEISHVILI, Svetlana; SEKERINA, Irina; MADUREIRA, Sandra. Verbal working memory assessment in Russian-Brazilian Portuguese bilinguals. *Cadernos de Linguística*, v. 2, n. 4, p. 01-24, e572, 2021. DOI 10.25189/2675-4916.2021.V2.N4.ID572
- SKRELIN, Pavel. German-Russian Language Contact: Is it in our power to foresee the flight of a word we have uttered? In: *GESUS-LINGUISTIK-TAGEN IN SANKT PETERSBURG*, 23., 2015. Hamburg, 2017. p. 65-74.
- SMIRNOVA HENRIQUES, Anna; FONTES, Mario A. de S.; SKRELIN, Pavel A.; KACHKOVSKAIA, Tatiana V.; RUSEISHVILI, Svetlana; BORREGO, Maria C.; PICCIN BERTELLI ZULETA, Patrícia; PICCOLOTTO FERREIRA, Léslie; MADUREIRA, Sandra. Russian immigrants in Brazil: to understand, to be understood. *Cadernos de Linguística*, v. 1, n. 2, p. 01-18, 2020. <https://doi.org/10.25189/2675-4916.2020.v1.n2.id210>
- SMIRNOVA HENRIQUES, Anna; RUSEISHVILI, Svetlana. Migrantes russófonos no Brasil no século XXI: perfis demográficos, caminhos de inserção e projetos migratórios. *Ponto-e-Vírgula*, 25, p. 83-96, 2019. <https://doi.org/10.23925/1982-4807.2019i25p83-96>
- SMIRNOVA HENRIQUES, Anna; SKOROBOGATOVA, Aleksandra S.; RUSEISHVILI, Svetlana; MADUREIRA, Sandra; SEKERINA, Irina A. Challenges in heritage language documentation: BraPoRus, spoken corpus of heritage Russian in Brazil. In: *INTERNATIONAL WORKSHOP ON DIGITAL LANGUAGE ARCHIVES*, 1., 2021. Proceedings [...]. Denton: University of North Texas, 2021. p. 22-24.
- SMIRNOVA HENRIQUES, Anna; SKOROBOGATOVA, Aleksandra S.; RUSEISHVILI, Svetlana; MADUREIRA, Sandra; SEKERINA, Irina A. BraPoRus, a Spoken Corpus of Elderly Heritage Russian in Brazil: Early Challenges and Future Plans. Presentation on the CLARE 5 (CORPORA FOR LANGUAGE AND AGING RESEARCH 5). Anchorage: University of Alaska Anchorage, 2022.
- SONIX. Available at <https://sonix.ai/>. Accessed on 29 January 2022.
- VOROBIEFF, Alexandre. Identidade e memória da comunidade russa na cidade de São Paulo. 244 p. Dissertation (Master's degree in Geography) - Departamento de Geografia, Faculdade de Filosofia, Letras e Ciências Humanas, Universidade de São Paulo, São Paulo, 2006. Available at <https://teses.usp.br/teses/disponiveis/8/8136/tde-18062007-141410/pt-br.php>. Accessed on 16 November, 2021.
- VOROBYEVA, Olga; ALESHKOVSKI, Ivan; GREBENYUK, Alexander. Russian Emigration at the Turn of the 21st Century. *Filosofia. Sociologia*, v. 29, n. 2, p. 107-118, 2018. DOI <https://doi.org/10.6001/fil-soc.v29i2.3706>.
- WULLF, Dirk U.; DAYNE, Simon de; JONES, Michael N.; MATA, Rui. The Aging Lexicon consortium. New perspectives on the aging lexicon. *Trends in Cognitive Science*, v. 23, n. 8, p. 686-698, 2019. DOI <https://doi.org/10.1016/j.tics.2019.05.003>.
- ZABOLOTSKY, Jacinto A. *A imigração Russa no Rio Grande do Sul: "os longos caminhos da esperança"*. Porto Alegre: Martins Livreiro, 1998.

## Información adicional

*Lay Summary:* The flow of Russophone migrants who searched for a place to start a new life in Brazil has intensified since the end of the 19th century. In the first half of the 20th century, Russia went through the Bolshevik Revolution and two world wars. After the Bolshevik Revolution of 1917, a number of Russophone migrants moved to Brazil, but most of them arrived in Brazil in the 1940s-1950s, after having lived in Europe or China, especially in Harbin. These migrants lived in communities and their children learned Russian in the family environment as their parents' heritage language. However, outside the home,

in most cases, they spoke Portuguese. Few went to Russian schools. Thus, the children spoke one language at home and another outside the home, and the older they grew, the less they used Russian, their heritage language. In the current work, we present strategies to study heritage Russian in Brazil within the scope of a project entitled BraPoRus that aims to analyze linguistic, behavioral and cultural aspects of the Russophone migrant community. To date, 26 heritage Russian speakers have participated in the survey, allowing us to record 167.5 hours of speech in Russian.