

Challenges in heritage language documentation: BraPoRus, spoken corpus of heritage Russian in Brazil

Anna Smirnova Henriques
LAEL, Pontifícia Universidade
Católica de São Paulo/PUC-SP
São Paulo, Brazil
anna.smirnova.liaac@gmail.com

Aleksandra S. Skorobogatova
FFLCH, Universidade de São
Paulo/USP
São Paulo, Brazil
as.skorobogatova@gmail.com

Svetlana Ruseishvili
Universidade de São Carlos/
UFSCar
São Carlos, Brazil
s.ruseishvili@gmail.com

Sandra Madureira
LAEL, Pontifícia Universidade
Católica/PUC-SP
São Paulo, Brazil
sandra.madureira.liaac@gmail.com

Irina A. Sekerina
College of Staten Island, The City
University of New York/CUNY
New York, USA
irina.sekerina@csi.cuny.edu

Abstract—The Bolshevik revolution in 1917, followed by the Civil War, induced a big wave of emigration from the ex-Russian Empire. These emigrants created their “Russia Abroad”. Many Russians stayed in Europe or China, but, in the 1940s and 1950s, many of them went to the USA, Latin America and other destinations. The importance of preserving the memories and documents of the old waves of the Russian emigration is crucial. Our group is collecting a corpus of heritage Russian in Brazil, the BRAZILIAN PORTUGUESE RUSSIAN CORPUS (BraPoRus). While the history of Russian immigration in Brazil is to some extent studied, their remarkably preserved Russian has not been described. Our current aim is to describe the BraPoRus, a corpus that consists of multiple speech samples of older Russian heritage speakers in Brazil, and to discuss the best ways to make these data available in the forms that satisfy the requirements both for the linguistic and sociological research.

Keywords—Russian as a heritage language; heritage speakers; language archives; oral history; bilingualism

I. INTRODUCTION

The Bolshevik revolution in 1917, followed by the Civil War, resulted a large emigration wave from the ex-Russian Empire: approximately 1.5 to 2 million Russians (this term frequently includes the Russophones or people who belong to “culturally russified communities”) left the country [1]. In their new countries, they founded churches, schools, local journals, and artistic groups, creating their “Russia Abroad”. Many Russians stayed in Europe [2], but the World War II forced them to emigrate as refugees to the USA, Latin America, and other destinations.

A special group of Russians is the “Russians from China”: most of them come from the village of Harbin which at the end of the 19th century began to receive Russian engineers in charge of the construction of the Chinese Eastern Railway, and later, after the Bolshevik Revolution and the arrival of many Russian immigrants, became an important destination for the Russian emigré community [3]. The Chinese Communist Revolution in 1949 brought important changes in the

immigration politics, and the majority of Russians left China in the 1950s, either moving back to the USSR or to the USA, Australia, and Brazil. Preserving the memories and documents of the old waves of the Russian emigration is crucial, but many archives were lost [1, 4]. It is also important to document and preserve speech samples of the Russian language spoken by these emigrants and their descendants. The digital collection of the Columbia University Libraries contains a number of oral history interviews with Russian immigrants, recorded in the 1960s [5], but this is a unique collection of such materials on this issue.

Brazil accepted the Russian refugees from the former Russian Empire between 1921 and 1941 [6]. In the post-World War II period, many Soviet Displaced Persons (DPs) and the families of the “white” Russian community in Europe also arrived to Brazil. In the 1950s, it was the turn of the Russian “white” stateless refugees from China [7]. Rough estimate based on the Brazilian Institute of Geography and Statistics 1950 census [8] is that 1,500 of these Russophone immigrants from China could be still alive. While the history of Russian immigration in Brazil is studied to some extent, their remarkably preserved Russian has not been described. Our group is collecting a corpus of moribund heritage Russian in Brazil, the BRAZILIAN PORTUGUESE RUSSIAN CORPUS (BraPoRus). D’Alessandro and colleagues [9] define a heritage language as moribund when it is spoken by elderly people who are a final generation of proficient speakers whose production and comprehension of heritage language are preserved; when they die, the language dies with them. Our current aim is to present the BraPoRus, a corpus that consists of 160 hours of speech samples of elderly Russian heritage speakers in Brazil ($M_{age} = 77$ years), and to discuss the best ways to make these data available in the forms that satisfy the requirements for the linguistic and sociological research in heritage language documentation.

II. PARTICIPANTS

The participants were selected according to the following criteria: 1) age 59 years and older (range: 59-98); 2) living in Brazil for most part of their life or being born in Brazil, speaking Portuguese in a natively way; 3) proficiency in Russian as a heritage language, sufficient to maintain a conversation for an hour; 4) no long-term residence in Russia; 5) no cognitive impairment. Currently, 31 participants (12 men and 19 women) are enrolled in the study.

III. METHOD

The protocol for data collection includes: 1) a brief demographic questionnaire; 2) a working memory test in Russian and Brazilian Portuguese using the Month-Ordering task [10]; 3) a semi-spontaneous narrative about the history of the participants' family and their immigration to Brazil; 4) the Bilingual Language Profile [11]; 5) a sociolinguistic interview with 139 questions adapted from the long HLVC (Heritage Language Variation and Change, Toronto) questionnaire [12]; 6) an assessment of narrative abilities in Russian and Brazilian Portuguese using Multilingual Assessment Instrument for Narratives (MAIN) [13]; 7) unscripted dialogues between participants in Russian; 8) intonation tasks; and 9) reading tasks. The data are being collected in 6-8 online sessions, through phone calls or videoconference by Zoom. All the speech samples recorded at the steps 3-7 will be transcribed and annotated using ELAN.

IV. RESULTS

The sociodemographic data that describe profiles of the BraPoRus participants are presented below.

TABLE I. SOCIODEMOGRAPHIC DATA

Age	
Mean	77.3 years (SD = 8.7)
Range	59-98
Residence	
São Paulo city	24
São Paulo state	2
Rio de Janeiro	4
Curitiba	1
Place of Birth	
China (Harbin)	13 (9)
Brazil	11
Europe	4
Russia	2
Belarus	1
Age of Arrival to Brazil (20 out of 31 participants)	
Median	10.8
Range	1-24

Only the two oldest participants (97 and 98) arrived in Brazil over the age of 18 (20 and 24, respectively), both born in Europe.

We have recorded the family history stories and, partially, sociolinguistic interviews in Russian from 21 participants. The total duration of these recordings currently is 160 hours. In addition, we have recorded 5.8 hours of speech in dialogue interactions.

V. DISCUSSION

In this project, we have recorded 160 hours of Russian speech samples produced by elderly Russian heritage speakers that reside in Brazil, mainly descendants of the emigrants that left the Russian Empire directly after the Bolshevik revolution. We plan to describe and characterize the bilingual speech of the elderly heritage Russian speakers. These will include attrition effects from Brazilian Portuguese, code-switching, intonation profiles, and interaction between the working memory and narrative abilities.

The sociolinguistic interviews collected in this study contain unique information about the history of Russian emigration, and the immigration experience after the Bolshevik revolution and World War II. Our questionnaires address the family history, the immigration paths, the places of residence and description of the houses and daily routine in Russophone communities of Europe, China and Brazil, the adaptation difficulties of Russophone families in Brazil, the life of the Russophone communities in Brazil, including education and cultural events, the religious traditions (some of the participants are from Old Believers families), food habits, and many other issues. In addition, some participants provided artifacts in the form of documents and old photos kept by the family.

As far as ethics is concerned, the common way of addressing ethical concerns in speech corpus is anonymization in ELAN and submission to an open database, such as the TalkBank [14]. The TalkBank Code of Ethics establishes that only age and location of the recordings could be annotated, but no personal data about individual participants. Many interviews contain sensitive data and should be anonymized; following the gold standards of the speech corpus construction, all the names and places mentioned in the recordings should be replaced with silence. From the other side, the names and places are necessary for the historical and sociological analysis, and the collections of oral history interviews, as in [5], contain these types of data.

VI. CONCLUSION

The rich data obtained in our project, which focuses on the study of the elderly Russian heritage speakers in Brazil, can be framed in two ways: as an annotated and anonymized speech corpus, and as a database of oral history interviews. This dualism raises many questions. Does it make sense to make the same interviews available in two forms, one as an anonymized corpus for linguistic research proposes and, at least with restricted access, as oral history interview database for use in history and sociology studies? If so, how could it be done? How to balance the participant anonymization and the preservation of the memories about the immigration history? How to guarantee that the linguistic data in specialized databases could be efficiently accessed by historians and sociologists in a useful format? The answers to these questions are important in order to improve the documentation of heritage languages and relate them to the historical and sociological studies of immigration.

ACKNOWLEDGMENT

Dr. Smirnova Henriques is supported by postdoctoral fellowship PNPd/CAPES (Programa Nacional de Pós-Doutorado da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior).

REFERENCES

- [1] M. Raeff, "Recent Perspectives on the History of the Russian Emigration (1920-1940)," *Kritika: Explorations in Russian and Eurasian History*, vol. 6 (2), pp. 319-334, 2005. DOI: <https://doi.org/10.1353/kri.2005.0025>.
- [2] A. J. Cohen, "Our Russian Passport": First World War Monuments, Transnational Commemoration, and the Russian Emigration in Europe, 1918-39," *Journal of Contemporary History*, vol. 49 (4), pp. 627-651, 2014. DOI: <https://doi.org/10.1177/0022009414538469>.
- [3] O. M. Bakich, "Emigre Identity: The Case of Harbin," *The South Atlantic Quarterly*, vol. 99 (1), pp. 51-73, 2000. Available at: <https://muse.jhu.edu/article/30644/>. Accessed on: 05 September 2021.
- [4] N. Saul, "American Collections on Immigrants and Émigrés from the Russian Empire," *Slavic & East European Information Resources*, vol. 4 (4), pp. 49-61, 2003. DOI: https://doi.org/10.1300/J167v04n04_05.
- [5] Columbia University Libraries, Digital Collections, Radio Liberty, 2021. Available at: https://dlc.library.columbia.edu/catalog?utf8=%E2%9C%93&search_field=all_text_teim&q=Radio%20Liberty. Accessed on: 05 September 2021.
- [6] A. Bytsenko, *Imigração da Rússia para o Brasil no início do século XX. Visões do paraíso e do inferno*. São Paulo: Universidade de São Paulo, 2006. Available at: <https://www.teses.usp.br/teses/disponiveis/8/8155/tde-12112007-132926/pt-br.php>. Accessed on: 05 September 2021.
- [7] S. Ruseishvili, *Ser russo em São Paulo. Os imigrantes russos e a reformulação de identidade após a Revolução Bolchevique de 1917*. São Paulo: Universidade de São Paulo, 2016. Available at: <https://teses.usp.br/teses/disponiveis/8/8132/tde-13022017-124015/pt-br.php>. Accessed on: 05 September 2021.
- [8] IBGE (Instituto Brasileiro de Geografia e Estatística), *Censo demográfico, 1956*. Available at: https://biblioteca.ibge.gov.br/visualizacao/periodicos/67/cd_1950_v1_br.pdf. Accessed on: 05 September 2021.
- [9] R. D'Alessandro, D. Natvig, M. T. Putnam, "Addressing challenges in formal research on moribund heritage languages: A path forward," *Frontiers in Psychology*, vol. 12, art. 700126, 2021. DOI: <https://doi.org/10.3389/fpsyg.2021.700126>.
- [10] D. Kempler, A. Almor, L. K. Tyler, E. S. Andersen, M. C. Macdonald, "Sentence comprehension deficits in Alzheimer's disease: a comparison of off-line vs. on-line sentence processing," *Brain and Language*, vol. 64 (3), pp. 297-316, 1998. DOI: <https://doi.org/10.1006/brln.1998.1980>.
- [11] D. Birdsong, L. M. Gertken, M. Amengual, *Bilingual Language Profile: An Easy-to-Use Instrument to Assess Bilingualism*, COERLL, University of Texas at Austin. Available at: <https://sites.la.utexas.edu/bilingual/>. Accessed on: 05 September 2021.
- [12] N. Nagy, "Heritage languages as new dialects," In: M. Côté, R. Knouihuizen, J. Nerbonne (Eds.), *The future of dialects*. Berlin: Language Science Press, 2016. DOI: <https://doi.org/10.17169/langsci.b81.81>.
- [13] N. Gagarina, D. Klop, I. M. Tsimpli, J. Walters, "Narrative abilities in bilingual children," *Applied Psycholinguistics*, vol. 37 (1), pp. 11-17, 2016. DOI: <https://doi.org/10.1017/S0142716415000399>.
- [14] B. MacWhinney, *The TalkBank system*, 2021. Available at: <https://www.talkbank.org/>. Accessed on: 05 September 2021.