

With approximately 1.5 million people, the Brazilian Polish community is considered to be the third largest Polish community in the world. Although the main waves of emigration to this country ceased shortly before the outbreak of the Second World War, more than 80 years ago, Brazil, and especially its southern region, remains a treasure trove of Polish linguistic and cultural traces. Some of these have been preserved in numerous polonica, among which a prominent place is held by the Polish-language press published in Brazil between 1892 and 1999, and especially until the late 1930s. Recognising more than a hundred periodicals published in Polish, including the oldest 'Gazeta Polska w Brazylii' (1892-1941) and the most popular and longest-edited weekly 'Lud' (1920-1999), as a source, still largely unexplored, of knowledge on the history of Polish immigrants and their descendants in Brazil, as well as their language, a detailed analysis is made of the Polish language recorded in their pages.

The source of the research material will be approximately 30,000-35,000 pages of material from the press of the Brazilian Polish community, which will be digitised and, after digital processing, made available in the form of a newly created text repository. Some of the press material we have collected has never been made available online, and until recently it was dispersed and deposited in Brazilian archives and religious archives, in Catholic parishes - mainly in and around Curitiba, the capital of the state of Paraná (southern Brazil). In this way, our planned activities fall within the scope of 'salvage linguistics', which involves the digital preservation of a rich collection of polonics, here of the press.

Our previous research on the language of the Brazilian Polish community has revealed that after more than 150 years of the 'separation' of the Brazilian Polish community from the Polish language in Poland, different phenomena have appeared in the language of the descendants of Polish emigrants than in the Polish language at home. In order to show them fully, we are planning an extensive linguistic analysis of the collected materials. It will cover the sociolinguistic, lexical, vocabulary, normative and cognitive planes. Sociolinguistic analyses will focus on the study of social patterns of language use and the role of Polish-language newspapers in the Polish community. Lexical research will focus on the description of Polish-speaking vocabulary, its vitality, dialectal affiliation and foreign language influences. Vocabulary research will include an analysis - including frequency analysis - of formants, derivation processes in the Polish language. On the other hand, normative and cognitive research will focus on the reconstruction of the linguistic image of the world of the descendants of Polish emigrants and the analysis of linguistic awareness and normative processes. Comprehensive research into the Polish language of the Brazilian Polish community will, we believe, allow us to reconstruct the way in which representatives of the descendants of Polish emigrants from the mid-nineteenth century who are still living in Brazil see the world, as well as to develop the norms of the Polish language of the Brazilian Polish community and its linguistic consciousness. Such a rich and multifaceted analysis has not yet been achieved by any Polish communicative community.

It should be emphasised that the research is planned to make very extensive use of digital humanities tools - the Corpusomat corpus tool and other digital tools, mainly CLARINs and the statistical cloud tool. In addition, as a result of the team's work, a publicly accessible searchable repository of these scans will be created; this repository will be deposited on a server purchased as part of the project and maintained (along with IT infrastructure) at the Faculty of Polish and Classical Philology of the A. Mickiewicz University in Poznań - the project applicant. In parallel with this - a corpus will be created in the Corpusomat tool, developed at the Institute of Computer Science of the Polish Academy of Sciences. This will create a unique collection of data, which will contribute to a better understanding of the evolution of the Polish language beyond Poland's borders. The project envisages publishing the results on an ongoing basis, which will make it possible to follow the progress of the research and use the collected repository by other researchers.