

Karolina Vyskocilova, Charles University in Prague

Preserving an Endangered Language Variety through a Specialized Corpus

This contribution will focus on Spoken Czech in Romanian Banat. A Czech language minority has been living in the South Eastern Romanian region of Banat since the first half of the nineteenth century and the community has been retained in the region until today. The language of this slowly disappearing language island has been captured in the Banat corpus, a continuously growing corpus which is a collection of spoken informal Czech used in Banatian villages and at the same time one of the biggest specialized corpora of spoken Czech.

In the first part, we will take a closer look at the corpus creation process – from the collection of the recordings through their processing and transcription (with a focus on the transcription method used) to the creation of the corpus. One of the great advantages of the corpus is the fact that it was built with a possibility of comparison in mind – it's practically fully comparable with the ORAL corpus, a corpus of spoken common Czech created by the Institute of Czech National Corpus. Therefore, both corpora can be used for comparison and verification of divergences between the two language varieties. The corpora could also assist with a re-examination of hypotheses from previous case studies which had to depend on knowledge, judgment and introspection of the linguists at the time the research was conducted.

The last part of the presentation will demonstrate the possibilities of the corpus on a few examples from Czech syntax. Firstly, it will examine enforced negation expressed by *nic* ['nothing'] and *nikerak* ['in no way'] which are typical features of Banatian Czech. Secondly, it will demonstrate the differences in the distribution and use of negative pronouns and adverbs (*nikde* [',nowhere'], *nikdo* [',nobody'] and *nikdy* [',never']) in both language varieties.

References

- Corpus BANAT*. Vyskočilová, Karolína. The Institute of the Czech National Corpus. Praha. URL: <<http://www.korpus.cz>>.
- Czech National Corpus - ORAL*. The Institute of the Czech National Corpus. Praha. URL: <<http://www.korpus.cz>>
- Gesce, Desideriu. 2013. *Historie českých komunit v Rumunsku* [History of Czech Communities in Romania]. Prague: Hermann & Synové [Hermann & Sons].
- Salzmann, Zdeněk. 2004. Bibliografie prací o rumunských Čechách a Slovácích [The bibliography of the studies about Romanian Czechs and Slovaks]. *Češi v Cizině* [Czechs Abroad] 12. 144–176.
- Utěšený, Slavomír. 1962. O jazyce českých osad na jihu rumunského Banátu [About the language in Czech villages in the south of Romanian Banat]. *Český Lid* [The Czech People] 49. 201–209.

- Utěšený, Slavomír. 1964. Z druhé výpravy za češtinou v rumunském Banátě [From the second expedition to Czech language in Romanian Banat]. *Český Lid* [The Czech People] 51. 27–32.
- Viková, Vilma. 1994. *Bígorská čeština: nástin jazykové monografie české vesnice v Rumunsku* [Bígr Czech: a sketch of a linguistic monograph of Czech villages in Romania]. Prague, CR: Faculty of Arts at Charles University dissertation.
- Vyskočilová, Karolína. 2012. *Syntaktická analýza projevů českých mluvčích v rumunském Banátu* [A syntactic analysis of Czech speakers in the Romanian Banat]. Prague, CR: Faculty of Arts at Charles University bachelor's thesis.
- Vyskočilová, Karolína. 2014. *Tvorba specializovaného korpusu banátské češtiny a jazyková analýza vybraných jevů*. [Building a specialised corpus of Banatian Czech with a linguistic analysis of selected features]. Prague, CR: Faculty of Arts at Charles University dissertation.