***Building the first digitalised learner corpus for Romanian (LECOR).***

***Difficulties and challenges***

Mihaela Cristescu, University of Bucharest

Carmen Mîrzea Vasile, University of Bucharest/ “Iorgu Iordan – Al. Rosetti” Institute of Linguistics, Romanian Academy

A learner corpus for Romanian is about to be built at the University of Bucharest, through the project *Learner Corpus of Romanian (LECOR). Collection, Annotation and Applications* (PN-III-P1-1.1-TE-2019-1066, funded by UEFISCDI, 2022-2024). The main goal of the LECOR project is to build and exploit the first digitalised learner corpus for Romanian, scalable and available in open-access format. The presentation will include general administrative and scientific data about the project and the current status of the project activities. A special attention will be paid to various types of difficulties encountered: difficulties during gathering the learners’ samples (written texts and spoken samples) and while recording learner- and task- variables (sociolinguistic information about the learner, the type of text and the circumstances in which it was produced); difficulties related to the morpho-syntactic and syntactic annotation (unclear text segments, errors falling into several categories, etc.); ethical and motivational issues. The presentation will contain various examples and will try also to show how profitable it is to have a corpus like LECOR for research and, finally, to improve teaching and learning Romanian as a foreign language.

***Élaboration de l’édition critique d’un dictionnaire roumain-romani du XIXème siècle***

Julieta Rotaru, INALCO

Aurore Tirard, INALCO

Nous présentons le travail qui a mené à l’édition d’une monographie sur les débuts de la lexicographie romani en Roumanie (Rotaru, Tirard et Shapoval 2022). Il s'agit de l'édition critique d’un dictionnaire roumain-romani rédigé dans les années 1870 par Vasile Pogor (1833-1906), descendant d'une ancienne famille aristocratique de Moldavie. Notre édition comprend une description grammaticale du matériel linguistique, des notices biographiques de l'auteur et une vaste bibliographie de ses œuvres. Toutes les entrées du dictionnaire roumain-romani ont été traduites en anglais. Deux dictionnaires inversés (romani-anglais et anglais-romani) ont été ajoutés en tant que dictionnaires des entrées examinées. Les auteurs du volume ont choisi de rejeter de nombreuses entrées copiées par l'auteur moldave à partir d'autres sources peu crédibles, et de maintenir certaines entrées douteuses, en les indiquant par un point d'interrogation. Sur le plan linguistique, l’élaboration de l’édition critique a posé un certain nombre de problèmes d’ordre graphématique et phonologique. Il a été très difficile de déterminer le dialecte décrit par le dictionnaire, en raison de la diversité des sources de l’auteur : données de terrain de première main, mais aussi copie d’autres auteurs ayant travaillé sur des dialectes très divers, tels que Grellmann (1783), de Réart (1835), de Rochas (1876), Vaillant (1844, 1868).

Grellmann, Heinrich Moritz Gottlieb. 1783. *Die Zigeuner, ein historischer Versuch über die Lebensart und Verfassung, Sitten und Schicksale dieses Volks in Europa, nebst ihrem Ursprunge*. Dessau, Leipzig : Verlags-Kasse.

Réart, Jaubert de. 1835. “Essai sur la langue des Bohémiens.” *Le publicateur des Pyrénées-Orientales*.

Rochas, Victor de. 1876. *Les Parias de France et d’Espagne (Cahots et Bohémiens)*. Paris : Librairie Hachette.

Rotaru Julieta, Aurore Tirard et Victor Shapoval. 2022. *Romani Lexicography in the Nineteenth Century. Volume 1: Lexicon Româno-Ţigănesc/Romanian-Gypsy Dictionary*. Munich : Lincom Europa.

Vaillant, Jean-Alexandre. 1844. *La Romanie ou histoire, langue, littérature, orographie, statistique des peuples de la langue d’Or, Ardialiens, Vallaques et Moldaves, résumés sous le nom de Romans*, vol. I. Paris : Arthus Bertrand.

Vaillant, Jean-Alexandre. 1868. *Grammaire, dialogues et vocabulaire de la langue des Bohémiens ou Cigains*. Paris : Typographie Pilloy.