Modelling Diatopic Variation in TEI: the Case of the VICAV Dictionaries

Donnerstag,
05.03.2020
14:15–14:45
VMP5 2101/2105

Karlheinz Mörth
Austrian Academy of Sciences
karlheinz.moerth@oeaw.ac.at

The *Vienna Corpus of Arabic Varieties* is a web-based project allowing researchers in the field of spoken Arabic varieties to collect and make available digital material concerning their field of interest, including both linguistically relevant data as well as methodological information with regard to data and tools potential be applied in digitally enabled dialectology. Irrespective of its name, VICAV has been working on a number of quite divergent types of digital language resources such as language profiles, linguistic feature lists, sample texts, bibliographies, dictionaries and documentation of digital tools and workflows. Being situated at the crossroads between diatopic linguistic approaches and research-driven text technology, the project has been serving quite diverse aims over time: teaching spoken varieties of modern Arabic, teaching comparatistic Arabic linguistics, teaching text encoding by means of the guidelines of the Text Encoding Initiative (TEI) as well as experimenting with new technologies.

The VICAV infrastructure has been used for different research projects so far, the two most important ones being TUNICO (*Linguistic dynamics in the Greater Tunis Area: a corpus-based approach*; FWF P-25706; 2013 – 2016) and TUNOCENT (*The linguistic terra incognita of Tunesia*; FWF P-31647; 2019 – 2023), both projects being focussed on contemporary spoken Tunisian Arabic. As part of the TUNICO project a digital corpus and a digital dictionary were created. The recently started TUNOCENT project will also create lexical data which is integrated into the existing TUNICO dictionary. While the TUNICO dictionary represented data encompassing diachronically distributed data, the new material contains data gleaned in several field campaigns from a large number of different locations. All VICAV digital language resources are encoded in TEI (P5). The presentation will give an outline of involved standards and discuss possibilities to model diatopic variation by means of the TEI's dictionary module.