



Researchers Launch Online Corpus of Palestinian Dialects



Photo by: Public Relations Office - Birzeit University

1 | 2 |
(#) (#)

An online corpus for the Palestinian dialect (called Curras) was launched on January 16, 2015 by the Computer Science Department, led by Professor Mustafa Jarrar.

Curras, which was funded by the Scientific Research Council of the Palestinian Education and Higher Education Ministry, is the first online archive of primary-source recordings of Palestinian dialects. It aims to fill the gap in applications that automatically annotate the Arabic dialect, specifically the Palestinian dialect, and keep pace with the dialectal content that is rapidly increasing on the internet, especially on social media.

Birzeit staff and student researchers collected texts in the dialectal form (about 56,000 words), such as posts from Twitter, scripts from the political spoof television show Watan Ala Watar, and others, annotating each word manually by describing its features (such as the part of speech, prefixes, suffixes, etc.).

The launch ceremony was attended by vice-chairman of the Higher Council for Scientific Research in Palestine, Professor Karim Tahboub, vice president of Academic Affairs Henry Giacaman, Faculty of Engineering and Technology dean Khaled Abaza, actors in Watan Ala Watar and a number of researchers and professors specialized in Arabic computational linguistics.

“Being involved in a range of projects in diverse digital humanities and having a wide-ranging technological engagement across the arts, humanities and cultural heritage sectors is very important nowadays, especially that the world is now dependent on online data,” said Dean Abaza.

In turn, Tahboub said that “Birzeit University has proven to be a hub for cutting-edge research and projects. It has always succeeded at being one of the most innovative universities in Palestine, and we at the Research Council of the Education and Higher Education Ministry are extremely honored to collaborate with the university and fund such projects that invest in our students and teachers’ capabilities to be world class researchers and become future leaders and innovators.”

Professor Jarrar, the project’s main researcher and principal investigator, emphasized the importance of building an annotated database for Palestinian dialects. Curras will create an online articulatory resource designed to improve understanding of how speech sounds are produced and provide Palestinians with a quick and easy tool for automatic translation, effective searching, and spell checking and speech recognition, among other things.

Jarrar expressed his appreciation for the support of external researchers who collaborated in the project, specifically Professor Nizar Habash from New York University at Abu Dhabi and Professor Owan Rambow from the University of Colombia in the U.S.

Curras is searchable online, and can be downloaded for free for research and academic purposes. For more information, visit the following website:

<http://portal.sina.birzeit.edu/curras> (<http://portal.sina.birzeit.edu/curras>).