Computational Processing of Arabic Dialects (invited talk)

Nizar Habash
Computer Science Department
New York University Abu Dhabi
nizar.habash@nyu.edu

1 Abstract

The Arabic language is a collection of variants among which Modern Standard Arabic (MSA) has a special status as the formal written standard of the media, culture and education across the Arab World. The other variants are informal spoken dialects that are the media of communication for daily life. As expected, most of the natural language processing resources and research in Arabic focuses on MSA. However, there is more work now on Arabic dialects, which includes efforts to build resources as well as exploit existing resources from MSA. In this talk, we present the challenges of processing Arabic dialects and the current advances in that area, with special emphasis on directions that exploit existing MSA (and other dialect) resources.

2 Author's Biography

Nizar Habash is an Associate Professor of Computer Science at New York University Abu Dhabi. He received his Ph.D. in 2003 from the Computer Science Department, University of Maryland College Park. He later joined Columbia University's Center for Computational Learning Systems where he co-founded in 2005 the Columbia Arabic Dialect Modeling group (CADIM) with Mona Diab and Owen Rambow. His research includes work on machine translation, morphological analysis, generation and disambiguation, and computational modeling of Arabic and its dialects. Professor Habash has over 100 publications including a book entitled "Introduction to Arabic Natural Language Processing". His website is http://www.nizarhabash.com.

Language Technology for Closely Related Languages and Language Variants (LT4CloseLang), page 1,
October 29, 2014, Doha, Qatar. ©2014 Association for Computational Linguistics