

Tutorial – May, 22nd 2006

Introduction to Arabic Natural Language Processing

The tutorial has four sections. First is a discussion of Arabic phonology and orthography with a focus on Arabic spelling peculiarities and their effect on Arabic processing. Arabic encoding issues are also addressed. Second, aspects of Arabic morphological phenomena are presented and explained. This is followed by a survey of different approaches to address these phenomena. Third, a survey of Arabic syntactic phenomena is presented and contrasted to English syntactic phenomena. Syntactic representation in the Arabic Treebank is discussed. Finally, Arabic dialects and the kind of problems they present for Arabic NLP are presented. Links to recent publications and available toolkits/resource for all four sections will be provided.

This tutorial is designed for computer scientists and linguistics alike. No knowledge of Arabic is required.

Outline:

40:00 Arabic Orthography

- Phonology
- Orthography
- Encoding issues

40:00 Arabic Morphology

- Introduction to Arabic Morphology
- Arabic Morphological Analysis/Generation

20:00 Coffee Break

40:00 Arabic Syntax

- Arabic Syntactic Phenomena
- Arabic Parsing Issues

40:00 Arabic Dialects

- Introduction to Arabic Dialects
- Processing of Arabic dialects

Instructor: Nizar Habash

Email Address: habash@cs.columbia.edu

Website: <http://www.nizarhabash.com>

Nizar Habash received his PhD in 2003 from the Computer Science Department, University of Maryland College Park. He is an Associate Research Scientist at the Center for Computational Learning Systems in Columbia University. His research focuses on Machine Translation, Natural Language Generation, Lexical Semantics, and Language Modeling. His work on Arabic ranges from research in Arabic encoding issues to Arabic-English Machine Translation and computational modeling of Arabic dialects for machine translation, speech recognition, and dialect parsing.