

MORPHOLOGICAL GENERATOR FOR URDU

Christopher Mala

M.S (Computational Linguistics)

IIIT-Hyderabad.

christopher.mpg08@research.iiit.ac.in

Abstract

Urdū, historically (spelled Ordu) is a Central Indo-Aryan language of the Indo-Iranian branch, belonging to the Indo-European family of languages. It is the official language of Pakistan and is predominantly spoken in Pakistan as well. Its vocabulary developed under Persian, Arabic and Turkic influence on Khariboli. It began to take shape during the Delhi Sultanate and Mughal Empire (AD 1526–1858) in South Asia.

Scholars independently categorize Urdu as a standardised register of Hindustani termed the standard dialect Khariboli. The grammatical description in this article concerns this standard Urdu. In general, the term "Urdu" can encompass dialects of Hindustani other than the standardised versions. The original language of the Mughals had been Turkish, but after their arrival in South Asia, they came to adopt Persian and later Urdu. The word Urdu is believed to be derived from the Turkish, word 'Ordu', which means army encampment. It was initially called Zaban-e-Ordu or language of the army and later just Urdu.. Due to recent progress in standardization of machine representation of text, applications such as machine translation and information retrieval are beginning to surface. Morphological analyzer and morphological generator are two essential and basic tools for building any language processing application for a natural language. Morphological analysis means to study the internal structure of the words of a language. A Morphological analyzer gives the morph analysis of a word i.e. for a given word a morphological analyzer will return its root word and word class along with other grammatical information depending upon its word class. For instance, it will provide gender, number, and case information for nouns and for verbs it will provide the information like tense, aspect, mood, etc. But the Morphological generator does exactly the reverse of it, i.e. for a given root word and its grammatical information it is capable of generating the word form of that root word. There is a need for a Morphological generator for Urdu that forms a vital integral part of applications like machine translation and universal dictionary. Here we present a tool that generates morphological forms of nouns, verbs, Adjectives, Adverbs and numbers of Urdu. Morphological Generator is developed based on the Morphological Analyzer which works on Word-Paradigm method and was developed by Prof. Mohd. Rehamath at DSS (HCU). The present Morphological Generator is developed in Perl® and will be made available with IL-ILMT system. This tool can be used as an engine.