ARABIC GENERATION IN THE FRAMEWORK OF THE UNIVERSAL NETWORKING LANGUAGE

Daoud Maher Daoud

ALzaytoonah University, Amman, Jordan
daoud_m@yahoo.com , daoud@maherinfo.com
&

GETA, CLIPS, IMAG
daoud.daoud@imag.fr

Abstract. This paper describes the work done on the developing of Arabic Deconversion within the framework of the Universal Networking Language (UNL). In this paper, the architecture of the system is explained along with the strategy used for the development. We also discuss issues and problems related to the UNL representation that affect the quality of generation. Additionally, the lingware engineering is introduced as a technique to enhance the quality and increase the development efficiency.

1   Introduction

Arabic is one of the world's main languages. It is the official language for over 289 million people. It is also the sacred language of nearly 1.48 billion Muslims throughout the world.

The alphabet consists of twenty-eight consonants but three of these are used as long vowels. Arabic also contains short vowel signs being indicated by marks above or below the letters. Like other Semitic languages, Arabic is written from right to left. It is a language characterized by rich morphology: most of the words are built from consonantal roots in which inflections and derivations are generated by vowel changes, insertions, and deletions.

The Universal Networking Language is a specification for the exchange of information. It is a formal language for symbolizing the sense of natural language sentences.

Currently, the UNL includes 16 languages. These include the six official languages of the United Nations (Arabic, Chinese, English, French, Russian and Spanish), in addition to ten other widely spoken languages (German, Hindi, Italian, Indonesian, Japanese, Latvian, Mongol, Portuguese, Swahili and Thai). In its second phase (1999 – 2005) the project will seek to further extend UNL access.

This paper presents the work completed on the generation of Arabic from UNL during the author's employment with Royal Scientific Society (RSS) in Jordan and his work on the UNL project. It described the work done on the generation of Arabic from UNL between 1996 till 1999. Since then, we think that the generation system maintained its main architecture.

Section 2 gives a description of the system that generates Arabic sentences from UNL representations. In section 3, we show how the system in implemented. In this perspective, issues such as mapping of relations, word ordering, and morphological generation are addressed. Results, future works, and conclusions are presented in Sections 4, 5 and 6 respectively.

2   THE ARCHITECTURE OF ARABIC GENERATION (DECONVERSION) SYSTEM

Universal networking language (UNL) is a semantic, language independent representation of a sentence that mediates between the enconversion (analysis) and deconversion (generation). It is a computer language aiming at removing language barriers from the Internet. The pivot paradigm is used: the representation of an utterance in the UNL interlingua is a hypergraph where normal nodes bear UWs ("Universal Words", or interlingual acceptions) with semantic attributes, and arcs bear semantic relations [13].

The sentence "Khaled bought a new car" can be expressed in UNL as:

agt(buy(icl>do(obj>thing),icl>purchase).@past.@entry, Khaled)