

Improving Adequacy in a Rule-Based English-to-Igala Automatic Translation System through Word Sense Disambiguation

Abstract

Here, we describe a rule-based machine translation system that translates English sentences to the Igala language. This work was evaluated with respect to adequacy, a measure of the quantity of information existent in the original text that the translated text contains. It indicates whether the output is a correct translation of the original sentence in a sense that the right meaning is communicated. An analysis revealed that adequacy was poor for sentences that contain words that exhibit ambiguity. The errors originated from the fact that correct meanings of ambiguous words were not selected based on the context. A further study of the architecture of the translation system showed it has no built-in module for word sense disambiguation. This accounts for the poor adequacy. A word-sense model was developed and incorporated into the MT system to disambiguate sentences before passing it to the translator. The model was implemented using JSP, a technology for building and deploying web applications in Java as the front end and MySQL as a backend. The model was tested on 100 previously translated sentences that contain ambiguous words. The level of adequacy increased from 32% to 68.2%.

Keywords: Translation, adequacy, rule-based machine translation system, word-sense disambiguation, ambiguity.

Introduction

Today, we live in an increasingly globalized and integrated world where **much** information is generated in various fields. However, since most of this information is in English, it remains out of reach to a large number of **people who do not speak the language**. As a consequence, **there is an** increasing demand for developing **a** means of translating from one language to another to enable efficient communication across cultures. Because human translation is expensive, time consuming and always in short supply, developing and deploying machine translation applications has become imperative for dealing with the problem of information inequalities created by multilingualism. Machine translation is the use of computers to automate some or all of the process of translating from one language to another [1]. Research organizations and government agencies now develop tools for automatic translation of text in order to achieve wider outreach and bridge the gap in language diversity [8].

Igala is a language in Nigeria. A significant percentage of Igala speakers cannot access information due to the fact information is mostly created in English. Developing **a** machine translation system that will efficiently translate English language to Igala to bridge this information gap has become a matter of absolute necessity in modern times. **Here, we describe a rule-based English-to-Igala Machine Translation System developed to achieve the objective of efficiently translating between these languages automatically.**

The human language is ambiguous because many words can be interpreted in multiple ways depending on the context. Ambiguity occurs when a word can have two or more different meanings or senses. Table 1 below illustrates this:

Table 1: Ambiguity illustrated

Word	Senses
plant	living/factory
tank	vehicle/container
poach	steal/boil
palm	tree/hand
axes	grind/tools
sake	benefit/drink
bass	fish/music
space	volume/outer
motion	legal/physical
crane	bird/machine
execute	Perform/kill
cold	Disease/temperature

Ambiguity leads to translation ambiguity which occurs when a word in the source language can be rendered in more ways in the target language [4]. Selecting the right meaning of an ambiguous word when performing translation is necessary to achieve accuracy of meanings. The task of assigning the correct or appropriate meaning to a given word in a text or discourse is called **word sense disambiguation** [7]. It is the computational identification of the meaning of words in context. The context of an ambiguous word is determined by the neighboring words.