A Rule Based Approach for Japanese-Uyghur Machine Translation System

Maimitili Nimaiti, Nagoya Institute of Technology, Nagoya, Japan
Yamamoto Izumi, Nagoya Institute of Technology, Nagoya, Japan

ABSTRACT

Japanese Uyghur machine translation system has been designed and developed using recent rule based approach. Even though Japanese and Uyghur language has many similarities, but there are also some linguistic differences cause serious problems to the word for word translation. In fact, as straightforward word-for-word Japanese-Uyghur translation sometimes yields unnatural Uyghur sentences. To raise the translation accuracy, the authors propose a word-for-word translation system using subject verb agreement in Uyghur. After a brief introduction to the comparative study of Japanese-Uyghur grammars, morphology and syntax, the authors explain their developing of a word to word rule base system. The coverage of this rule base system, the rules for translation, comparison of experimental result between statistical machine translation system and rule base machine translation system are explained. Some practical suffix translation methods solving problems in Uyghur language are also proposed.

Keywords: Japanese, Machine Translation, Suffix Translation, Uyghur, Word for Word Translation

1. INTRODUCTION

Uyghur is spoken by 8.5 million (2004) in China, mostly in the far western Xinjiang Autonomous Region. The necessity of machine translation system is growing rapidly due to the increase of government documents and translation request. One of them is the lack of training data in both automatic speech recognition and machine translation(MT), especially for low-resource languages.
Significant progress has been made by various research groups towards the goal of getting reliable statistical translation results. Researchers focus their efforts on enhancing the way different tasks of MT are performed. There are very few researches on MT in Uighur. They are implemented between English to Uighur MT system and Japanese to Uighur MT system uses the rule based approach, whose all knowledge from linguists is externalized as a set of inference rules. In these work, a translation system is implemented that works on word by word translation And case suffixes are considered only for both languages (Muhtar, 2001; Muhtar, 1994; Yasuhiro, 1997). Actually Both Japanese and Uighur include lots of suffixes. The harmonization about Uighur language is not explained clearly. These approaches have several drawbacks related to time consumption and rule conflict. In this work we pay much attention to linguistic information than translating approaches.

Subject-verb agreement is very important for Uyghur language processing. Because there is a big difference between Japanese and Uighur in expressing the grammatical category of person and number of a noun, and in verb forms which require some affixes to express different tense of an action in a sentence. In this paper, we specifically examine the function of a subject verb approach for translation result and emphasize the effectiveness of this function.

2. GRAMMATICAL COMPARISONS OF JAPANESE AND UIGHUR

Uyghur, like all the other Turkic languages, has a word order of subject+object+verb (SOV), and is considered to be an agglutinative language with very productive inflectional and derivational suffixation process in which a sequence of inflectional and derivational morphemes get affixed to a word stem. In Uyghur, a verb could have hundreds of word forms by sequentially adding different affixes to the word stem. Japanese, which is also considered to be an agglutinative language, also has the same word order and morphological features as Uyghur. Some researches show that this morphological and syntactic closeness is sufficient to obtain a relatively good translation result from Japanese into Uyghur on a transfer approach (Yasuhiro Ogawa, 2000; Polat Kadir, 2004). In the following sections, we will make a comparison between Japanese and Uyghur in two different levels: morphology and syntax with a close attention focused on their differences.

2.1. Morphological Comparison

As we compare the word formation, we could find that in both Japanese and Uyghur, word forms are generated by attaching many suffixes denoting case, mood, person, tense, etc. to one word stem as seen in Example (1).

(1) kuralmiganliktin ("as it was not seen")
kur + al + mi + ghan + liqtin
(見られなかったので)
kur/看出:stem
+al/れ:passive voice
+mi/な:negation
+ghan/かった:past tense
+liqtin/の
d: causal form

Generally, Japanese and Uyghur share a significant amount of morphological and syntactic features in common. However, there are also some differences
Generic Cabling of Intelligent Buildings Based on Ant Colony Algorithm
[www.igi-global.com/article/generic-cabling-intelligent-buildings-based/55128?camid=4v1a](www.igi-global.com/article/generic-cabling-intelligent-buildings-based/55128?camid=4v1a)

Search for an Optimal Solution to Vague Traffic Problems Using the PSK Method
[www.igi-global.com/chapter/search-for-an-optimal-solution-to-vague-traffic-problems-using-the-psk-method/207206?camid=4v1a](www.igi-global.com/chapter/search-for-an-optimal-solution-to-vague-traffic-problems-using-the-psk-method/207206?camid=4v1a)