ABSTRACT

The rule based approach to machine translation (MT) confines grammatical rules between the source and the target language with the goal of constructing grammatical translation between the language pair. In this paper, we describe the structural representation of English stemmer, POS tagging and design transfer rules which can generate Hindi sentence from the structural representation of the English sentence. Due to the specific terminology of homoeopathic sentences and the linguistic gap between the two languages the translation of these literatures from English to Hindi is a challenging task. The rule sets are used to plug gap between the two languages. Further, rule sets are described for mapping preposition verbs, nouns, etc. Finally, a system architecture has been proposed for the translation of homoeopathy literature from English to Hindi Language. The system accuracy has been evaluated using Bleu score, which is found out to be 0.7501 and the accuracy percentage of the system is 82.23%.

Keywords: Agreement Rules, Corpus, Homoeopathy Machine Translation, PoS Tagging, Stemmer

INTRODUCTION

Machine translation (MT) has been defined as the process that utilizes computer software to translate text from one natural language to another. Generally, for translating the source language (SL) to target language (TL) different approaches/techniques are used. Some of the major approaches are Rule based, Statistical approach and Example based approach. The MT system described in this paper, uses rule based approach, which is one of the effective techniques (Vaishali & Devale, 2010).

In recent years, the research in MT has gained momentum. Many language pairs (including the languages of Indian origin) have been chosen to develop MT systems, with good results. In the work of Jain, Bahadur and Chauhan(2012), the target language generation mechanism has been outlined for English to Sanskrit language pair using rule based machine translation technique for their system “Etrans”. In adopting rules based MT from English to Bangla, Francisca, Mamun Mia and Rahman (2011, Jun-Jul) have proposed language translation model that relies on rule based
methodologies particularly fuzzy rules. Most of these systems are in the English to Hindi languages (specific domain) with the exceptions of a Hindi to English (Sinha, & Thakur, 2005) and English to Kannada (Kumar, & Murthy, 2006) machine translation systems. Indian machine translation systems are used to translate English to Hindi language (Naskar & Bandyopadhyay, 2002), which are as follows; The English to Hindi Anusaaraka system follows the basic principles (Bharti, Chaitanya, Kulkarni, & Sangal, 1997) of information preservation. The system makes text in one Indian language accessible in another Indian language. Anubharti 2004 (Sinha, 2004) approach for machine-aided-translation is a hybridized example-based machine translation approach that is a combination of example-based, corpus-based approaches and some elementary grammatical analysis.

In this work, we describe an English-Hindi MT system for Homoeopathy documents. In our proposed system, sentences of Homoeopathy literature (in English) are first passed to Stemmer (Dwivedi, & Sukhadeve, 2011, May). Further, the Part of Speech (PoS) tagger (Dwivedi, & Sukhadeve, 2011, July), tags all the words with the help of our designed tagging rules and corpuses (English and Hindi). These steps are assisted by English to Hindi online dictionary Shabdkosh (शब्दकोश). Afterwards, we compare the structural representation of both the language for mapping and designed assembly rules for translation.

**Structural Representation of Homoeopathic Sentences**

Our description begins with basic information about word order and agreement principles, to facilitate the translation of English sentences to Hindi. In the translation process, we observed that there are differences between English and Hindi languages for sentences. The early work in this area for English has been carried by Quirk, Greenbaum, Lееch, and Svartvik (1985) and Allen (1995), and that for Hindi some significant contributions are made by Bharati, Chaitanya, and Sangal (1995) and Sastri and Apte (1968).

Our further study of Homoeopathy literatures(Dhama & Dhama, 2008), reveals that many times its sentences follow different representations and structures which are normally not observed in other medical languages. To explain this, we took three different types of Homoeopathic sentences.

1. “To combat high fever; Pyrogenium 200, \(\frac{1}{2}\) hourly (3).”
2. “Swelling very red with shining and burning; fever with dry hot skin; and worse in warm room.”
3. “Ring-shaped eruptions; much itching and lemon coloured crust, remedy is Tellurium 30 or 200; 10 mins. (3).”

In sentence (a): - The structure of the sentence is different because before acquiring the meaning of complete sentence, a semicolon is used which interrupts the meaning of the sentence. This type of sentence is difficult to tag grammatically. The structure of the sentence, “S ADJ ; OBJECT (Num, Adj)” is different from the normal (allopathic) English language sentence of the similar structure, “In fever take Paracetamol 500 mg, twice a day”. The structure of this sentence is “Subject(S) Verb (V) Object (O)”. (Dwivedi, & Sukhadeve, 2011, July)

In sentence (b):- There are three parts in a single sentence separated by semicolon (;) and the meaning of the sentence is not complete. The structure of the sentence would be, “S ADJ OBJECT ; S ADJ OBJECT ; Conj S ADJ OBJECT”. Similar sentence in the other medical language would be “The patient is dull, heavy and stupefied look.” The structure of this sentence is “S V O”.

Sentence (c): - The sentence has a different style specially used in homeopathy (doctor’s prescription) having different potencies with single remedy. The structure of this sentence would be “S; ADV V Conj S OBJECT; S V O (Num Conj Num; Num N Num)”. In allopath, however, similar sentence would take a simple structure like “For severe pains take one tablet daily of Buta 250 mg for three days.” The structure of this sentence is “S V O”.

Copyright © 2012, IGI Global. Copying or distributing in print or electronic forms without written permission of IGI Global is prohibited.
Forward Dendritic Spikes: A Mechanism for Parallel Processing in Dendritic Subunits and Shifting Output Codes
Oscar Herreras, Julia Makarova and José Manuel Ibarz (2009). Advancing Artificial Intelligence through Biological Process Applications (pp. 42-59).
www.igi-global.com/chapter/forward-dendritic-spikes/4971?camid=4v1a