

TranSish: Translator from Sanskrit to English-A Rule based Machine Translation

Pankaj Upadhyay^{Å*}, Umesh Chandra Jaiswal^Å, Kumar Ashish^Å

^ÅComputer Science and Engineering Department, Madan Mohan Malviya University of technology, Gorakhpur (U.P) India

Accepted 09 Aug 2014, Available online 10 Oct 2014, Vol.4, No.5 (Oct 2014)

Abstract

Sanskrit is the primary liturgical language of Hinduism and a scholarly literary language that is in the use of lingua franca in the Indian cultural zone. English is a West Germanic language that was first spoken in early medieval England and is now a global lingua franca. There is a lot of English words of Sanskrit origin For example, English mother is a cognate of Sanskrit *mātār'*; likewise, father-pitār, brother-bhrātār, sister-svāsār, son-sūnū, daughter-duhitār, man-manu/manav, dental-dāntam, nose-nas, cow-gāuḥ etc. Machine translation from Sanskrit to English is a suitable phenomenon to understand the Paninian grammar corresponding to English grammar as translation being the most important applications of Natural Language Processing. With the help of Artificial Intelligence we provide an interface named transish a Translator from Sanskrit to English that convert Sanskrit sentences to English with a rule based model of parser and a semantic mapper.

Keywords: Transish, Sanskrit Machine Translation, Parser, Paninian grammar, Rule-Based Model.

1. Introduction

Machine translation can use a method based on linguistic rules that is translated in a linguistic way or a process that utilizes software to translate source language (here Sanskrit sentences) into target language (English). Rule-based methods parse a text, usually creating an intermediary, symbolic representation, from which the text in the target language is generated. Transish (translator from Sanskrit to English) an interface of web based phenomenon is used here for translation that follow the rule based approach. The human translation process may be described as:

1. Decoding the meaning of the source text;
2. Re-encoding this meaning in the target language.

Behind this ostensibly simple procedure lies a complex cognitive operation. To decode the meaning of the source text in its entirety, the translator must interpret and analyse all the features of the text, a process that requires in-depth knowledge of the grammar, semantics, the text, syntax, idioms, etc., of the source language, as well as the culture of its speakers. The translator needs the same in-depth knowledge to re-encode the meaning in the target language. Transish uses parser and transliteration process for translation. The robustness of RBMT (rule based machine translation) is high and works on exact match reasoning that is used by transish.

2. Literature review

Sanskrit, as defined by Pāṇini evolved out of the earlier Vedic form. The beginning of Vedic Sanskrit can be

traced as early as 1700–1200 BCE. Scholars often distinguish Vedic Sanskrit and Classical or Pāṇinian Sanskrit as separate dialects. Though they are quite similar, they differ in a number of essential points of phonology, vocabulary, grammar and syntax. In the 2001 census of India, 14,135 people reported Sanskrit as their native language. Since the 1990s movements to spread spoken Sanskrit have been increasing organizations like *Samskrita Bharati* conduct Speak Sanskrit workshops to popularize the language. In India, Sanskrit is among the 14 original languages of the Eighth Schedule to the Constitution. The state of Uttarakhand in India has ruled Sanskrit as its second official language. According to Prajapati (2005), more than 3000 Sanskrit works have been composed in the period post Indian Independence (i.e., since 1947) alone.

The Sahitya Akademi has had, since 1967, an award for the best creative work written that year in Sanskrit. In 2009, Satyavrat Shastri became the first Sanskrit author to win the Jnanpith Award, India's highest literary award. In the Republic of India, in Nepal and Indonesia, Sanskrit phrases are widely used as mottoes for various national, educational and social organisations for example: Republic of India: 'Satyameva Jayate'. Means: Truth alone triumphs. Nepal: 'Janani Janmabhūmisca Sv argadapi gariyasi'. Means: Mother and motherland are greater than heaven.

3. Unicode representation

Unicode is a computing industry standard for encoding, representation and handling of text expressed in most of the world's writing systems. The Unicode Consortium was

*Corresponding author: Pankaj Upadhyay

incorporated on January 3, 1991, in California, and in October 1991, the first volume of the Unicode standard was published. Unicode can be implemented by different character encodings. The most commonly used encodings are UTF-8, UTF-16 and the now-obsolete UCS-2.

UTF-8 uses one byte for any ASCII character, all of which have the same code values in both UTF-8 and ASCII encoding, and up to four bytes for other characters. In the transish project UTF-8 is used in order to translation from Sanskrit to English. UTF-8 is an 8-bit variable-width encoding which maximizes compatibility with ASCII.

4. Design of proposed model

In this machine translation Sanskrit sentence as a source language is input through an interface called transish interface. Second step is to preprocess the Sanskrit text and then parse it into corresponding character with the help of Unicode to corresponding English characters then apply a transish algorithm to get the final output as English Sentences as:

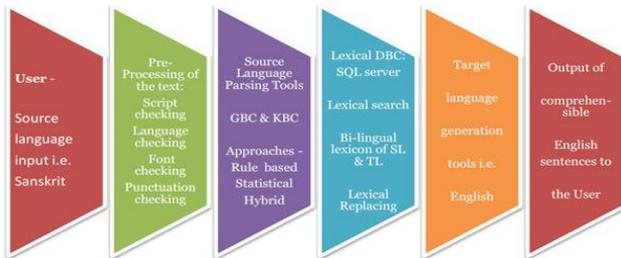


Fig.1 Design of translator from Sanskrit to English

5. Rule based Machine Translation

Major approaches of Machine Translation are rule-based machine translation (RBMT, also known as the rational approach). Rule based translation consists of:

- (1). Process of analyzing input sentence of a source language syntactically and or semantically
- (2). Process of generating output sentence of a target language based on internal structure each process is controlled by the dictionary and the rules.

There are three different types of rule-based machine translation systems:

Direct Systems (Dictionary Based Machine Translation) map input to output with basic rules.

Transfer RBMT Systems (Transfer Based Machine Translation) employ morphological and syntactical analysis.

Interlingua RBMT Systems (Interlingua) use an abstract meaning. Transish used direct system for translation of Sanskrit to English.

6. Approach for designing Transish algorithm

The transish algorithm take input as Sanskrit sentence and follow the given approach.

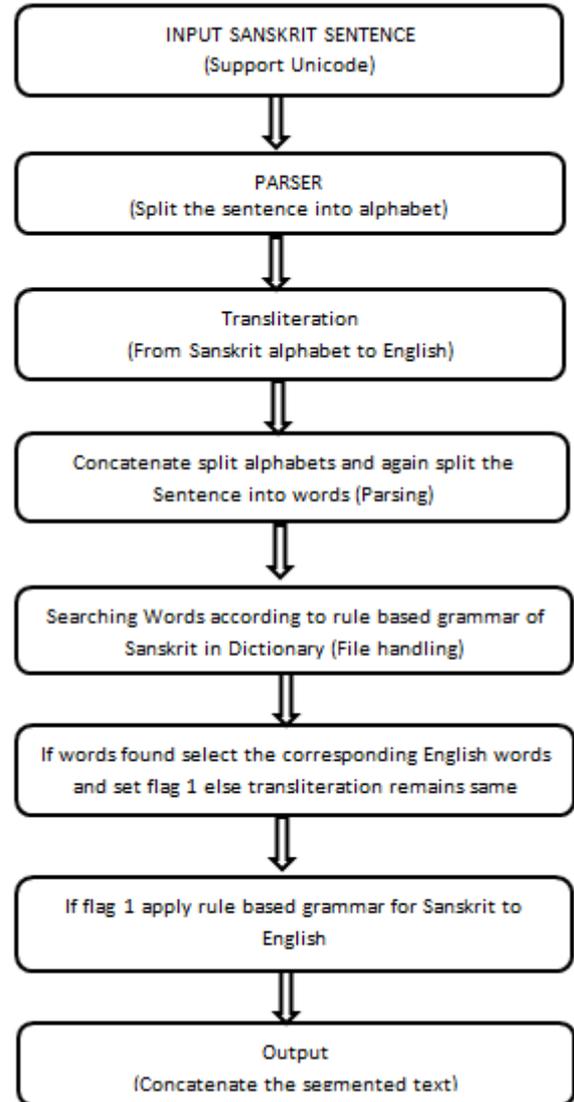


Fig.2 Approach for designing Transish algorithm.

There are six tenses in Sanskrit, out of this six transish first focus on Lat lakāra (present tense) to process the Sanskrit sentences for English translation. Details of Lat lakāra (present tense) parasmaipada is given below.

Table 1 Lat lakāra (present tense) parasmaipada

Person	Singular Number	Dual Number	Plural Number
First Person	Ti (ति)	tasतस् (तः)	Anti (अन्ति)
Second Person	Si (सि)	Thasथस् (थः)	Th (थ)
Third Person	Miमि	Vasवस् (वः)	Mas मस् (मः)

7. Transish Algorithm

1. Take the input from the text area inputted by virtual Keyboard of Unicode and set the flag=0;
- \$_SESSION ['x'] ← \$_POST ['txtarea'];
3. Check out translate button is set on any event occurs:

The front-end then assigns phonetic transcriptions to each word, and divides and marks the text into prosodic units, like phrases, clauses, and sentences. The process of assigning phonetic transcriptions to words is called *text-to-phoneme* or *grapheme-to-phoneme* conversion. Phonetic transcriptions and prosody information together make up the symbolic linguistic representation that is output by the front-end.

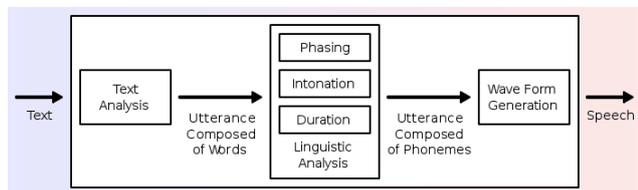


Fig.6 Typical TTS System



Fig.6a text to speech for corresponding output as inputted by the user at run time based on TTS System.

9.4 Accuracy level: Accuracy of transish is measured by options chosen by user. If user analyses that transish translation is excellent, select 5 and according to the user satisfaction transish process it.

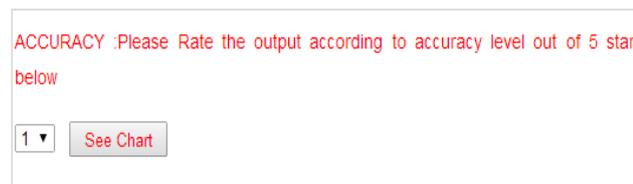


Fig.7 Select options for rating accuracy.

9.5. Transish graph: PHP combined with MySQL are cross-platform. After rating by the user it can fetch the current data and add the user rating to its corresponding data and update the database to show the updated results.

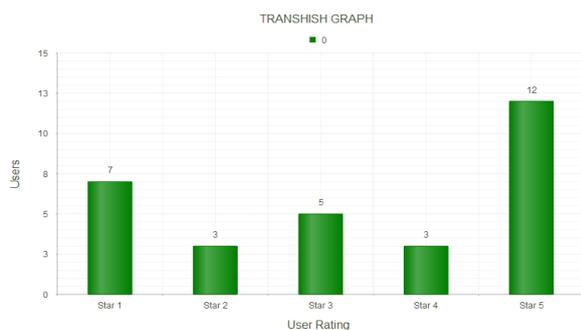


Fig.8 User point of Transish graph on user satisfaction.

Conclusions and Future scope

In this proposed paper we have mainly focused On Transish for Lat lakāra (present tense) parasmaipada as Compared with tenses of Sanskrit and as translation done a text to speech engine process the output and provide the facility to play the outputted sentence. As per knowledge a translator from Sanskrit to English is not available or may be under process. Transish help the people to interact with the Sanskrit language in easy way. The Sanskrit has no order that is a free world order language. Thus we get each word separately and maintain a separate meaning after applying sandhi rule and other grammar to understand it deeply. As translation is not so far in this area so future scope would be broad but need time and practice and the most important is the deeply knowledge of Sanskrit. One can extend the transish for other tenses and extend the transish web based thesis. The basic objective of the paper is to make the rule based transish a translator from Sanskrit to English and the proposed algorithm make the transish more efficient and user friendly.

References

(April-2013) Sanskrit is second official language in Uttarakhand – the Hindustan times. Hindustantimes.com. 19 January 2010 retrieved 2012-04-05, IJART, Volume 2, Issue

Vimal Mishra and r. B. Mishra, Study of example based English to Sanskrit machine translation, http://polibits.Gelbukh.com/37_05.pdf.

Sandhi splitter and analyzer for Sanskrit Sachin sanskrit.jnu.ac.in/rstudents /mphl / Sachin.pdf

Kumar, Sachin, GirishNathJha (2005) a Paninian sandhi analyzer for Sanskrit in the souvenir abstracts of platinum jubilee international conference of the Linguistic Society of India (2013), university of Hyderabad, 36-37 Sins (print): 2319 – 2526, volume-2, issue-2

Learn Sanskrit in 30 days anandsp1.files.wordpress.Com/ 2011/08/learn-sanskrit-in-30-days.pdf

A practical Sanskrit introductory <ftp://ftp.nac.ac.za/Wikner/sktintro.ps600-letter-march97> <tp://ftp.nac.ac.za/wikner/sktintro.ps600-a4paper-march97>

Wikipedia for Sanskrit grammar and wiki for TTS and lucent of Sanskrit grammar

Van Santen, Jan p. H. Sproat, Richard w olive, joseph p Hirschberg, Julia (1997). *progress in speech synthesis*, Springer ISBN 0-387-94701-9.