A Sanskrit to Hindi Language Machine Translator using Rule Based Method

Prateek Agrawal
Lovely Professional University,
Punjab, India.
Univeristy of Klagenfurt, Austria
`prateek061186@gmail.com`

Vishu Madaan
Lovely Professional University,
Punjab, India.
`vishumadaan123@gmail.com`

1 Demonstration

Hindi and Sanskrit both the languages are having the same script i.e. Devnagari Script which results in few basic similarities in their grammar rules. As we know that Hindi ranks fourth in terms of speaker's size in the world and over 60 Million people in India are Hindi internet users [1]. In India itself, there are approximately 120 languages and 240 mother tongues but hardly a few languages are recognized worldwide while the others are losing their existence in society day by day. Likewise, Sanskrit is one among those important languages that are being ignored in society. As per census report of India in 2001, less than 15000 citizens have returned Sanskrit as their Mother tongue or preferred medium of communication. A key reason behind poor acceptance of Sanskrit is due to language barrier among Indian masses and lack of knowledge about this language among people. Therefore, our attempt is just to connect a big crowd of Hindi users with Sanskrit language and make them familiar at least with the basics of Sanskrit. We developed a translation tool that parses Sanskrit words (prose) one by one and translate it into equivalent Hindi language in step by step manner:

- We created a strong Hindi-Sanskrit corpus that can deal with Sanskrit words effectively and efficiently (Agrawal and Jain, 2019).

- We proposed an algorithm to stem Sanskrit word that chops off the starts / ends of words to find the root words in the form of nouns and verbs (Jain and Agrawal, 2015).

- After stemming, we developed an algorithm to search the equivalent Hindi meaning of stemmed words from the corpus based on semantic analysis (Bhadwal et al., 2019).

- We developed an algorithm to implement semantic analysis to translate words that helps the tool to identify required parameter details like gender, number, case etc.

- Next, we developed an algorithm for discourse integration to disjoin each translated sentence based on subject / noun dependency (Bhadwal et al., 2020).

- Next, we implemented pragmatic analysis algorithm that ensures the meaningful validation of these translated Hindi sentences syntactically and semantically (Agrawal and Jain, 2019).

- We further extended our work to summarize the translated text story and suggest the suitable heading / title. For this, we referred ripple down rule-based parts of speech (RDR-POS) Tagger (Dat et al., 2016) for word tagging in the POS tagger corpora.

- We proposed a title generation algorithm which suggests some suitable title of translated text (Jain and Agrawal, 2018).

- Finally, we assembled all phases to one translation tool that takes a story of maximum one hundred words as input and translates it into equivalent Hindi language.

Figure 1 shows the working demonstration of proposed translation tool in nine window panels and 18 widgets (objects) to describe each step. In the first window panel, 3 objects are highlighted and numbered as 1, 4 and 5. Users can type any Sanskrit sentence using a simple QWERTY keyboard in the text area (highlighted as Object 1).

---

[1] http://www.business-standard.com/article/current-affairs/hindi-internet-users-estimated-at-60-million-in-india-survey-116020400922$_1$.html
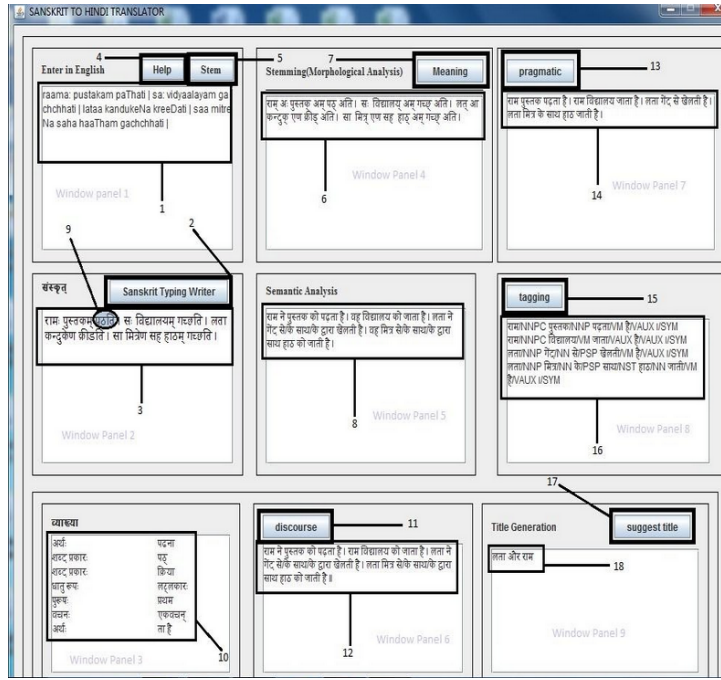
Figure 1: Screenshot of Output Interface to translate Sanskrit sentence into Hindi Language

Help button (as Object 4), is added to help novice users to type Sanskrit sentence(s) in English. Simultaneously, transliterated sentence from English to Sanskrit is displayed in the second window panel as object 3. Sentences can also be typed using a virtual keyboard (highlighted as object 2) [6]. Next step is to stem Sanskrit words written in the 2nd window panel by pressing. Stem button (highlighted as object 5 in first window panel). Output of the same is displayed in the 4th window panel and highlighted as object 6. Next step is to press the button "Meaning" (Object 7) to get Hindi meaning of stemmed Sanskrit words (highlighted as Object 8 in window panel 5).

Once we get raw or intermediate meaning of Sanskrit words into Hindi as object 8, we click on the Discourse button (as object 11) to get the translated results in discourse form (highlighted as object 12 in window panel 6). As an option for easy understanding, our system has been felicitated to explain the meaning of Sanskrit words into Hindi with complete explanation (highlighted as object 9 and 10). Once Discourse integration is done, the next step is to get the correct meaning of sentences (as object 14) by clicking on the pragmatic button (as object 13) in window panel 7. Final translated sentences are passed through RDR POS tagger for tagging purposes (as highlighted in window panel 8 using objects 15 and 16). Finally, the system implements an algorithm to generate appropriate

title(s) of translated sentences in Hindi using object button 17 in Window panel 9 and displays the result as object 18.

References

Prateek Agrawal and Leena Jain. 2019. Anuvaadika: Implementation of sanskrit to hindi translation tool using rule-based approach. *Recent Patents on Computer Science*, 12(2):10–21.

Neha Bhadwal, Prateek Agrawal, and Vishu Madaan. 2019. Bilingual machine translation system between hindi and sanskrit languages. In *International Conference on Advanced Informatics for Computing Research*, pages 312–321. https://link.springer.com/chapter/10.1007/978-981-15-0108-1_29.

Neha Bhadwal, Prateek Agrawal, and Vishu Madaan. 2020. A machine translation system from hindi to sanskrit language using rule based approach. *Scalable Computing: Practice and Experience*, 21(3):543–553.

Quoc Nguyen Dat, Quoc Nguyen Dai, Duc Pham Dang, and Bao Pham Son. 2016. A robust transformation-based learning approach using ripple down rules for part-of-speech tagging. *AI Communications (AICom)*, 29(3):409–422.

Leena Jain and Prateek Agrawal. 2015. Text independent root word identification in hindi language using natural language processing. *International Journal of Advanced Intelligence Paradigms*, 7(3/4):240.

Leena Jain and Prateek Agrawal. 2018. Sheershak: an automatic title generation tool for hindi short stories. In *2018 International Conference on Advances in Computing, Communication Control and Networking (ICACCCN)*, pages 579–584. IEEE.