International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING
ISSN:2147-6799 www.ijisae.org Original Research Paper

Importance of Sanskrit Language in Natural Language Processing and Machine Translation: a Review

Vaishali Kherdekar*[1], Sachin Naik[2], Dhananjay S. Deshpande[3], Pravin Metkewar[4]

Submitted: 26/01/2024 **Revised**: 04/03/2024 **Accepted**: 12/03/2024

Abstract: This paper presents an importance of the Sanskrit language in Computer Science and Information Technology i.e. how the Sanskrit language plays a key role in the research area. This paper also presents how the Sanskrit language is important in various fields of computer science such as natural language processing, automatic speech recognition, machine translation, etc. It is also beneficial for human health. In Sanskrit, sentences can be formed using a minimum number of words than in any other language. In this paper reasons for selecting Sanskrit as a Natural Language Processing are presented. This paper also presents the importance of the Sanskrit language in research.

Keywords: *Information Technology, Machine Translation, Natural Language Processing, Panini's grammar, Sanskrit.*

1. Introduction

Sanskrit is the mother language of all Indo-European languages. Most of the books such as Upanishads, Vedas, Smriti, Shruti, Puranas, Ramayana and Mahabharata are present in the Sanskrit language.[2]

Computer Science and Information Technology have an important aspect in the creation of new computing algorithms and machine that has powerful cognitive and computational abilities. Sanskrit is the most powerful language because of its grammar. It is lexically [5] productive language. The grammar has simple, effective rules for conversion between various forms and sentences. The vocabulary of Sanskrit language consists of maximum number of words than any other language in the world. It has the power to say a sentence in a minimum number of words than any other language. Sanskrit is the best computer-friendly language. The Sanskrit language can be very useful if it will be used in computers and technology.

This paper covers the following points.

1. Importance of Sanskrit language in NLP

2. Research Areas of Sanskrit language.

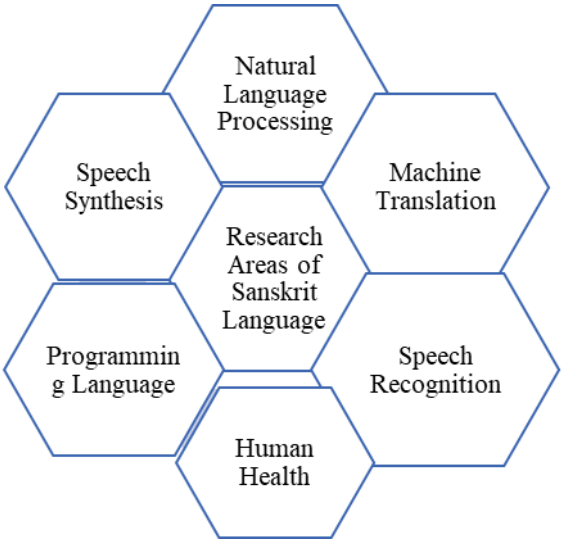

Fig 1. Research Areas of the Sanskrit Language

2. Literature Review

It was stated in Forbes [13] that Sanskrit is the best computer-friendly language because of the large number of vocabularies present in it than any other language. Learning Sanskrit improves brain functioning. After learning Sanskrit students get better marks in other subjects i.e. it enhances memory power. The Sanskrit language plays an important role in human health by increasing energy levels and reducing stress. The energy level of the human body increases by stimulating the body points with the help of reading, speaking, or reciting. All nerves of the tongue are used in Sanskrit which improves blood circulation, mind relaxation, and enhancement of brain functioning. Bhalodia,

[1] *MIT Art's Commerce and Science College, Alandi(D), Pune, Maharashtra State, India.*
ORCID ID: 0000-0003-3291-5080
[2] *Symbiosis Institute of Computer Studies and Research (SICSR), Symbiosis International (Deemed University) (SIU), Pune, Maharashtra State, India.*
ORCID ID: 0000-0001-9502-4133
[3]*MBAESG , Ajeenkya D Y Patil University, Pune, Maharashtra State, India.*
ORCID ID: 0000-0001-7131-1584
[4]*Dr Vishwanath Karad MIT World Peace University, Pune, Maharashtra State, India.*
ORCID ID: 0000-0003-4079-0011
** Corresponding Author Email: vaishali.kherdekar@gmail.com*

Table 1. Existing Literature of the Sanskrit Language in various areas

Ref No.	Research Work	Research Area	Author	Features
Jan-16	Use of Sanskrit for natural language processing	Natural Language Processing	Chandana Bathulapalli et al.	Due to nature of unambiguity, it is useful for NLP.
Feb-13	Sanskrit as a Programming Language and Natural Language Processing	Natural Language Processing	Shashank Saxena and Raghav Agrawal	Build dependency parser using DFA.
Mar-14	Natural Language Processing and Sanskrit	Natural Language Processing	Deeptanshu Jha et al.	Build Parseer and Semantic Analyzer.
Apr-21	Automatic Speech Recognition in Sanskrit: A New Speech Corpus and Modelling Insights	Automatic Speech Recognition	Devaraja Adiga et al.	Segmentation based on vowel split gives better result.
May-11	Challenges in Developing a TTS for Sanskrit	Speech Synthesis	Diwakar Mishra et al.	Linguistic analysis of word and sentence recognition is challenging.
Jun-15	Study of Methodologies for utilizing Sanskrit in Computational linguistics	Programming Language	Prajakta R. Chaudhari et al.	Useful as a computational programming language
Sep-15	Sanskrit as a Programming Language: Possibilities & difficulties	Programming Language	Vipin Mishra	It is perfect in morphology, semantics, and its structure.
Oct-16	Stop-Word Removal Algorithm and its Implementation for the Sanskrit Language	Natural Language Processing	Jaideepsinh K. Raulji and Jatinderkumar R. Saini	Stop word removal algorithm is designed for text analytics.
14-2020	A Review on various approaches in Machine Translation for the Sanskrit Language	Machine Translation	Santosh Deshpande and Neha Kulkarni	Due to its morphological nature it is challenging in machine translation.
15-2019	General Structure of Sanskrit Machine Translation System	Machine Translation	K. S. Gilda et al.	Machine translation is challenging because of several meaning of words.

16-2020	Improving Neural Machine Translation for Sanskrit-English	Machine Translation	Ravneet Punia et al.	Developed Sanskrit to English machine translation system using NN.
17-2010	ANN and Rule Based Model for English to Sanskrit Machine Translation	Machine Translation	Vimal Mishra and R.B.Mishra	English sentences are translated into Sanskrit sentences using ANN

Jitendra V in his paper [3] stated that Sanskrit is best language for computer because processing can be done fast in it. The program can be executed from two sides means meaning of sentence does not change in Sanskrit when changing the words to be executed(processed). Existing literature reveals that the correct pronunciation of Sanskrit words improves the speech tone and also helps in speech therapy. Rashmi Jha and et.al. in their paper [8] talked about Charles Babbage statement about Panini's Grammar that "The structure of Paninian Grammar is nothing but a computer program". It has captured the base of universal principles of all languages. Rick Briggs stated that a minimum number of words used to transfer a message to a computer system using Sanskrit.[7] C-DAC developed a computer application program called "Pandu-lipi" Samshodaka has browse, search, index, analyze and hyper linking features. In [15] author developed a Sanskrit Machine Translator System module which consists of components such as morphological analyzer, language converter, Sandhi analyzer, Vibhakti analyzer, Karaka analyzer etc. In [16] author developed a system that translates Sanskrit words into their equivalent English words using a neural network based on reinforcement learning and transfer learning. In the future, there is a lot of scopes to work on it. In[17] author designed a rule base machine translation system using Artificial Neural Network to convert English sentences into their equivalent Sanskrit sentences. In future author wants to use case based reasoning model and data mining technique for more accurate results. In [18] author designed a Parser system that splits Upsarg and Pratyay of Sanskrit words using Panini's Sanskrit grammar framework. In the future author wants to design a parser that will parse the Sanskrit sentences by considering Vibhakti, Sandhi, Samas and Lakar with higher accuracy.

3. Fact about Sanskrit

These An important fact about the Sanskrit language is Panini's grammar. It consists of rules to form each word in Sanskrit. It mainly focuses on Semantic rules, roots, vocabulary categorization, categorization of suffices, and gender studies.

Sanskrit is the most technological and organized language in the world. It is an innovative language. It is one of the most suitable languages for computers. It is a very efficient language for making algorithms. It is lexically productive [4] language

Sanskrit is a highly versatile language because it has a lot of synonyms for a single word. It is one of the most accurate languages in pronunciation.

3.1 Features of the Sanskrit Language:

A unique feature of the Sanskrit language is a dual case [3] which is not present in any other language. The Sanskrit language has intelligent grammar and words describing the [3] properties rather than objects. Special attention to the dual case, extensive use of Vibhaktis, support for high inflection, and refined pronunciation of each and every word. To make the meaning of a sentence independent based on the position of its constituent words, Sanskrit has one more feature called as inflection-based syntax.

3.2 Importance of Sanskrit on Human Health

Phonetics of Sanskrit has roots in energy points of the body and reading and speaking Sanskrit stimulates these points and raises the energy levels which increases the resistance power against illness, relaxation to the mind, and reduces stress.

Sanskrit is the only language, which uses all the nerves of the tongue. By its pronunciations, energy points in the body get activated which causes blood circulation to improve which helps in controlling blood pressure, diabetes, and cholesterol.

Sanskrit is proven to be helpful in Speech Therapy. Research shows that learning the language improves brain functioning and boosts memory power and concentration.

Due to its importance in technology and in human health, Sanskrit is accepted all over the world. In Germany 14, Australia 8, the USA 3, UK 2 of the top universities teach Sanskrit.

NASA too has a department in it to research Sanskrit manuscripts. It is a highly regularized language and suitable

for computers. [10] In fact, NASA declared that Sanskrit is the only “Unambiguous Spoken language on the planet”.

To learn Ayurveda, understanding of Sanskrit language is very important because the science of Ayurveda is written in the Sanskrit language.

3.3 Importance of Sanskrit in Computer:

In [1] author reported that due to well-structured grammar and unambiguity, the Sanskrit language is useful for processing. NASA scientists also believed that Sanskrit is useful in the development of artificial intelligence applications. In [2] author developed a parser using the rules of the Sanskrit language present in book Ashtadhayayi on Sanskrit grammer. In [3] author reported the reason of selecting Sanskrit as a Natural Language Processing, and also noted that Sanskrit can be useful for machine speech therapy. In [4] automatic speech recognition is developed by creating 78 hours of the dataset. In [5] ‘Samvachak’ a Sanskrit text to speech.

4. 4. Natural Language Processing

Natural Language Processing is the part of artificial intelligence in which machines are built that understand written or speech data just like humans. It has two phases natural language understanding and natural language generation. There are various applications of NLP such as human-computer interface, speech synthesis, speech translation, etc. Most natural languages are ambiguous, having more than one meaning. There are various types of ambiguities present such as scope ambiguity, attachment ambiguity, semantic ambiguity, and pragmatic ambiguity.

Sanskrit is the best language for NLP due to the following reasons.

1. Sanskrit is more conducive to be approximated to a semantic net model used in artificial intelligence due to the rich and rigid set of nouns indicating their relationship to each other.[1]

2. Due to the presence of Vibhakti’s in Sanskrit, it conveys the correct, precise, specific, and unambiguous meaning of the sentence. [1]

3. Sanskrit language helps us to generate NLP because words describe the properties.[6]

4. IT support for high inflection and attention is on the dual case.[3]

5. Each and every word is refined and pronounced properly.[3][6]

6. Avoid misconception, and mispronunciation, and increasing the viability of language.[6]

7. Presence of more than 4000 rules and Sutras such as Paribhasha, Samjna, Adhikara etc.

5. Machine Translation

Machine translation is the application [14] of Natural Language Processing. To share ideas and transfer information there is a need for inter-language translation. With the help of machine translation one natural language can be decoded into other. Due to the presence of different types of ambiguities, it is very challenging. The Sanskrit language overcome the problem of ambiguity in natural language processing.

Sanskrit is the best language for Machine Translation due to the following reasons.

1. Sanskrit language has the knowledge of linguistic sounds. It is called Panini Shikha Shastra.

2. In the generation and utilization of the words from stems, the Sanskrit language has meaningful parts of words. It is known as ‘Pada Vyutpatti'.

3. There is one-to-one correspondence in every lexical unit of the Sanskrit language.

4. Due to the presence of Vibhakti, there is an absence of ambiguity in the Sanskrit language.

5. In "Vakyapadiyam” a wonderful Vyakaran is present in Sanskrit which represents the meaning of words with respect to context.

6. Challenges

Systems for the processing of speech in the Sanskrit language do not exist. Tokens in Sanskrit text are very long and recognizing longer sequences due to Sandhi and compounding is very challenging. [3] Due to the presence of ‘Sandhi’ and ‘Samasa’ Sanskrit language has a large number of tokens in text.[4] Linguistic analysis for word and sentence recognition is challenging for Sanskrit language. [5]

7. Conclusion

Although Sanskrit is an ancient language, due to its features it is useful in various applications. This paper presented the importance of the Sanskrit language in research and on human health. Based on the literature we have concluded that there is scope to work on Sanskrit in various research areas such as natural language processing, automatic speech recognition, speech synthesis, machine translation etc.

Acknowledgements

Researchers thank the faculty of Symbiosis Institute of Computer Studies & Research Centre for providing support and guidance.

Author contributions

All the authors have equal contribution towards this study.

Conflicts of interest

The authors declare no conflicts of interest.

References

[1] Bathulapalli, C., Desai, D., & Kanhere, M. (2016). Use of Sanskrit for natural language processing. Int. J. Sanskrit Res, 2(6), 78-81.

[2] Saxena, S., & Agrawal, R. (2013). Sanskrit as a programming language and natural language processing. Global Journal of Management and Business Studies, 3(10), 1135-1142.

[3] Jha, D., Jha, R., & Varshney, V. (2014). Natural Language Processing and Sanskrit. International Journal of Computer Engineering & Technology, 5(10), 57-63.

[4] Adiga, D., Kumar, R., Krishna, A., Jyothi, P., Ramakrishnan, G., & Goyal, P. (2021). Automatic speech recognition in Sanskrit: a new speech corpus and modelling insights. arXiv preprint arXiv:2106.05852.

[5] Mishra, D., Jha, G. N., & Bali, K. (2011, March). Challenges in Developing a TTS for Sanskrit. In International Conference on Information Systems for Indian Languages (pp. 228-231). Springer, Berlin, Heidelberg.

[6] Chaudhari, P. R., Gangurde, P. C., & Kulkarni, N. L. (2015). Study of methodologies for utilizing Sanskrit in computational linguistics. International Journal of Electronics, Communication and Soft Computing Science & Engineering (IJECSCSE), 1.

[7] Briggs, R. (1985). Knowledge representation in Sanskrit and artificial intelligence. AI magazine, 6(1), 32-32.

[8] Jha, R., Jha, A., Jha, D., & Jha, S. (2016, March). Is Sanskrit the most suitable language for natural language processing?. In 2016 3rd International Conference on Computing for Sustainable Global Development (INDIACom) (pp. 211-216). IEEE.

[9] Mishra, V. (2015). Sanskrit as a Programming Language: Possibilities & Difficulties. International Journal of Innovative Science, Engineering & Technology, 2(4).

[10] Nair, R. R., & Devi, L. S. (2011). Sanskrit Informatics: Informatics for Sanskrit studies and research. Centre for Informatics Research and Development. Book

[11] Raulji, J. K., & Saini, J. R. (2016). Stop-word removal algorithm and its implementation for Sanskrit language. International Journal of Computer Applications, 150(2), 15-17.

[12] Sanyala, K., & Paula, P. (2021). A Study of aindra School of Sanskrit Grammar in the Light of Paninian Framework in Natural Language Processing. Multicultural Education, 7(11).

[13] Chakraborty, S. (2021). The Role of Specific Grammar for Interpretation in Sanskrit. Journal of Research in Humanities and Social Science, 9 (2), 107-187.

[14] Deshpande, S., & Kulkarni, M. N. (2020). A Review on various approaches in Machine Translation for Sanskrit. International Journal of Future Generation Communication and Networking, 13(2), 113-120.

[15] Gilda, K. S., Dixit, S. S., & Narote, S. V. (2019). General Structure of Sanskrit Machine Translation System.

[16] Punia, R., Sharma, A., Pruthi, S., & Jain, M. (2020, December). Improving Neural Machine Translation for Sanskrit-English. In Proceedings of the 17th International Conference on Natural Language Processing (ICON) (pp. 234-238).

[17] Mishra, V., & Mishra, R. B. (2009). ANN and Rule Based Model for English to Sanskrit Machine Translation. INFOCOMP Journal of Computer Science, 9(1), 80–89.

[18] Panchal, B., Parmar, A., Dholitar, S. (2019). An Approach of Splitting Upsarg and Pratyay of Sanskrit Word Using Paninian Framework of Sanskrit Grammar. In: Abraham, A., Dutta, P., Mandal, J., Bhattacharya, A., Dutta, S. (eds) Emerging Technologies in Data Mining and Information Security. Advances in Intelligent Systems and Computing, vol 814. Springer, Singapore.