



INDIAN KNOWLEDGE SYSTEMS DIVISION

Ministry of Education, Govt of India.

AICTE Headquarters, Nelson Mandela Marg, Vasant Kunj, New Delhi-110070. Website:
www.iksindia.org *Tel: 011-29581523/1004.*

PROJECT COMPLETION REPORT

Sanskrit word order following the steps of target language

SUBMITTED BY

Amba Kulkarni

Department of Sanskrit Studies

University of Hyderabad

Date of Submission: 24th Dec 2024

FUNDED BY

**Indian Knowledge Systems Division of Ministry of
Education, Govt of India**

IKS Division File Reference No: AICTE/IKS/RPF-1/2021-22/01

Project Duration: From May 2022 to Oct 2023

1. Project Title: Sanskrit word order following the steps of target language

2. Duration of Project: From May '22 To Oct '23

3. Focus Area: Computational Linguistics with focus on Machine Translation

4. Principal Investigator Name: Amba Kulkarni

Organisation/Institution: University of Hyderabad

5. Co-Principal Investigator Name: Pavankumar Satuluri

Organisation/Institution: IIT Roorkee

6. Collaborating Organisations. Not Applicable

Part A - Summary Report

1. Project Objectives

Objective as per the approved Project	Fully Achieved/ Partially Achieved (indicate shortfall)	Reason for Partial Achievement
To develop computational modules that will change the word order in a given Sanskrit shloka so as to suit the native word order of the target language.	Fully Achieved	
To develop this anvayakrama generator for the following 5 languages: Marathi, Telugu, English, German and French	Partially achieved	During the review meetings we were advised to work only on the Indian languages, and hence we did not work on the Foreign languages.

2. Deliverables

Deliverables as per the approved Project	Fully Achieved/ Partially Achieved/ Not achieved	Reason for Partial Achievement
Parallel corpus of around 1,000 shlokas and their anvaya following Hindi, Marathi and Telugu word order	Fully Achieved	

Deliverables as per the approved Project	Fully Achieved/ Partially Achieved/ Not achieved	Reason for Partial Achievement
Anvaya Generation programme	Fully achieved	

3. Specific Benefits/Outcome

- Patent, if any —NIL
- Product/Process developed/ technology transferred

The modules are integrated with the Sanskrit Computational Linguistics Toolkit Samsaadhanii. It is available at the following URLS

— <https://sanskrit.uohyd.ac.in/scl>

— <https://scl.samsaadhanii.in/scl>

- List of Publications arising from the Project
 - A journal paper is being prepared
- Linkages developed — NIL

4. Manpower trained

- No. of Students based on the project: One
- No. of Master's Thesis/Reports based on the project:

The project fellow Mr Sanal Vikram is about to submit his PhD thesis on this topic.

- Position and Number of Other Technical Staff Trained: Nil

5. Methodology (Minimum 250 Words):

The current trend in NLP is to use the Machine Learning approaches. However, we decided to follow the rule based and knowledge based approach. The main reason behind this is the following.

Indian civilization has a unique contribution in the form of language sciences. The systematic study of a language from various dimensions starting from phonetics to semantics through grammar, etymology, etc. and sustaining through 25 centuries is the asset we have. In order to nurture this further, and not only to preserve the tradition, but allow it to grow further, we would like to explore the knowledge that is hidden in these scriptures. This will also help us to rejuvenate the studies of language sciences deeply rooted in Indian thoughts.

In Indian traditional learning books, the **anvayakrama** is explained in the following verse, to convert a poetry into prose order:

विशेषणं पुरस्कृत्य विशेष्यं तदनंतरम् |

कर्तृ-कर्म-क्रिया-युक्तं एतदन्वयलक्षणम् || (Samāsachakram -10)

That is, starting with the adjectives, targeting the headword, in the order of *kartr-karma-kriyā* (subject-object-verb) gives an *anvaya*.

We used these insights together with the Heuristic data to generate an anvayakrama for Sanskrit as well as the target languages - Marathi, Hindi and Telugu.

- **Salient Research Achievements**

- Demonstrating the usability of rule based methods for Machine Translation

- **New Observations (minimum 200 words):**

- We noticed that while the rules given in the traditional shastras are useful, they are not sufficient, and we had to come up with the heuristic rules to complement them. For example, the shastras are silent on where the kaala-adhikaranas and desh-adhikaranas be placed. Also they are silent about the placement of various avyayas, etc.
- Several divergences between Sanskrit and modern Indian languages have been observed with respect to the order of the words in these languages.

- **Innovations(Give bulleted list):**

- A very simple method of traversing the dependency tree following the word order of the target language as specified in their grammars results in the anvaya order according to the target language.

- Use of Panini's grammar, and especially the kaaraka and non kaaraka relations play an important role.
- Ranking various syntactico-semantic relations using the Heuristics
- Use of Machine Learning approaches to arrive at the ranking automatically

6. Indication of scope for future work (Maximum 300 Words):

- a) The current programmes can be extended to all the Indian languages, with a little bit of twicking.
- b) The modules are also independent, and can be plugged in very easily in any MT system.
- c) The programmes could also be further improved so as to work on the given input using the information at morphological level so as to produce some output when the parser fails.

7. Procurement/Usage of Equipment:

Sr	Name of Equipment	Make/Model	Cost (Rs)	Date of Purchase & Installation	Remarks regarding maintenance/ breakdown
1	Laptop	Dell	64,068	29th Sept 2022	In Good condition

Part B – Comprehensive Report

1. Project Title : Sanskrit word order following the steps of target language

2. Literature Review

Sanskrit is inflectionally rich and it rarely uses position to encode any syntactic or semantic relation between words. This enables Sanskrit to move the words freely in a sentence. Within Indian tradition, the word order was regarded as free, provided the proximity (sannidhi) is not violated. The Indian theorists found the sentences with different arrangements of words to be equivalent in meaning, with an exception of subject-predicate (uddes'ya-vidheya). The difficulty in understanding the verses due to deviation from the 'default' word order, however, had been realised. The commentators commenting on the verses have followed this 'default' word order known as the anvaya of a verse (s'loka)¹. But this default word order is not followed strictly either by the commentators or by the authors while composing prose. When the Sanskrit texts started being translated into fixed word order European languages, the free word order of Sanskrit had been noted. The westerners tried to propose a framework for the Sanskrit word order. However, these studies also lead to only a preferential, or frequent arrangement, and not 'the' arrangement. The deviations were considered to be the exceptions. Even while discussing the syntax, much emphasis had been laid on the concord and government rather than order.

Though Sanskrit allows free movement of words, and there are preferential word orders, certainly not all permutations of the words are allowed. So the attention of the researchers then was drawn to the restrictions on word order rather than possible word orders. The first systematic account of the word order in Sanskrit was by [Staal \(1967\)](#). He introduced a concept of 'wild tree' which allows the movement of the items within a phrase freely. In this model, any two items from different phrases

cannot interleave or the words in one phrase cannot leave their place and move to the domain of other phrases.

This model was formalised and empirically tested by Gillon on a corpus of approximately thousand Classical Sanskrit prose sentences. Half of the sentences of this corpus was from a single text,² and the other half was selected at random from the sentences found in Apte (1925). His empirical observations confirm the model suggested by Staal with an exception of movement of genitives and adjectives across the clause boundaries.

Aralikatti (1991) studies the modern Sanskrit texts and conversations from the point of view of the flexibility in word order. He found that the modern writings and the conversations follow the default word order. Scharf et al. (2015) presents the preliminary observations with regards to the comparison of the word order in prose and verse which confirm more flexibility in verses than in prose.

There was no computational model that can generate a prose word order from the verse. This is the first attempt towards that direction.

Our research concentrates on the generation of the prose word order taking insights from the traditional knowledge. This provides an opportunity to demonstrate the suitability of the Indic Knowledge Systems in the domain of Language Technology.

3. Methodology adopted for Development of the Product/ Process in the Research Project (*State briefly in 800 words*)

Authors who have written Sanskrit prose also have taken advantage of the free word order to present texts that are consistent with the intended meter or are interesting from the aesthetics point of view. But it is also true that it is difficult to understand poetry compared to prose. This is evident from the fact, we notice, that the commentators, especially commenting on the kāvyā (poetic) literature, first rewrite the verse in prose in some default word order, and then comment on it. This deviation from the normal word order adds an extra load on the part of the readers in understanding the poetry. In order to understand such texts, one needs special training for interpreting these texts. We come across commentaries on several of such Sanskrit poetic texts, which make their understanding easier.

In the Indian tradition, we see two methods followed by commentators while dealing with sentence level analysis of śloka (Tubb and Boose, 2007). In both these approaches, the aim of the commentator is to unfold the encoded meaning. While doing so, the commentator takes clues from the theories of śābdabodha. The two approaches are described below.

- The first approach is known as Khaṇḍa-anvaya (also known as katham-bhūtinī), where the commentator starts with the verb, and the expectancies associated with the verb, and goes on filling these slots with the nominal forms in the śloka. Once the basic skeleton with all the expectancies is ready, then the commentator connects the viśeṣaṇas (adjectives) to their viśeṣyas (headwords), providing flesh to the skeleton.

The parse produced by the machine provides us the khaṇḍānvaya. All the words that are directly related to the verb work as a backbone, or as a part of the sentence carrying core information. The adjectives attached to the nouns, the arguments of non-finite verbs, etc. typically occupy the second or higher level in the tree structure, and add the flesh to the structure.

- The second approach is the Daṇḍa-anvaya (also known as anvaya-mukhī). In this method, first the commentator arranges the words in the śloka in a prose form, following a default word order typically encountered in prose.

Initializing Reordering Task

Anvaya reordering tool is a simple script written in Python. It takes the set of quintuplets as input and creates a corresponding Python tree object. Since multiple morphological variants of a word cannot occur in a single set of dependency solution, variant information is not used presently but is preserved for proposed uses in the future.

Graphical representation of the tree object created with the parsed information of the Bhagavadgītā verse is same as in Figure 5, without the dotted lines.

Deciding the Order

We found the clues for anvaya-order in the Samāsacakra. The two relevant kārīkās go like this.

Ādau kartṛpadam vācyam dvitīyādīpadam tataḥ Ktvātumunlyap ca madhye tu kuryād ante kriyāpadam (Samāsacakram kārīkā 4, (Bhagirath, 1901, p. 12))

Starting with kartṛ, followed by other words, placing the non-finite verbal forms such as ktvā, tumun, lyap in between, place the main verb at the end.

Viśeṣaṇam puraskṛtya viśeṣyam tadanantaram Kartṛ-karma-kriyā-yuktam etad anvaya-lakṣaṇam (Samāsacakram kārīkā 10, (Bhagirath, 1901, p. 13))

Starting with adjectives, targeting the headword, in the order of kartṛ-karma-kriyā (subject- object-verb), gives an anvaya (the natural order of words in a sentence).

In recent studies, Aralikatti (1991) has shown that the unmarked word order in Sanskrit is SOV. That is, all the arguments of a verb are placed to the left of the verb starting with the kartṛ, then karman followed by other arguments, the attributive adjectives are placed to the left of the noun they qualify, and the predicate is at the end of the sentence. The sub-ordinate clauses, if any, are before the predicate.

Taking clue from these resources, we define a sentence to be in canonical word order if it satisfies the following criteria:

All the modifiers are placed to the left of the word they modify.

This is equivalent to the following.

1. The adjectives are to the left of the substantives they qualify.
2. All the arguments of a verb (either in finite form or in non-finite form) are to its left.

3. All the non-finite forms that modify the finite verb form are to its left.

This implies that the main verb⁷ is always the last word of a sentence. This canonical word order provides us the Daṇḍa-anvaya for ślokas. We assigned the priorities to the dependency relation labels following these clues. These priorities were further fine-tuned by studying the commentaries and prose orders of around 400 ślokas from literature including Bhagavadgītā, Nītiśataka, various subhāṣitas and about 50 poetic prose sentences from Kādambarī.

Adjusted by various measures, currently, the relative positions of various arguments are fixed following the rules given below.

1. Sambodhya (vocative) comes at the initial position in the canonical order.
2. Karṭṛ comes after vocative.
3. Kāraka relations follow in reverse order i.e. adhikaraṇa, apādāna, sampradāna, karaṇa and karman.
4. Viśeṣanas, modifiers with genitive case markers, etc. are placed before their viśeṣya.
5. Kriyāviśeṣana, pratiṣedha etc. are placed immediately before their corresponding verb.
6. Mukhyakriyā is positioned at the end of the sentence.
7. Avyaya particles such as tu and api are placed immediately after their parent word.
8. The non-finite verbal forms are placed before the karman. All the arguments of non-finite verb appear to their left.
9. The karṭṛ-samānādhikaraṇa and karma-samānādhikaraṇa are placed after the karṭṛ and karman respectively.

Sorting the Tree

The reordering tool traverses through the tree object using level-order-iteration and sort recursively at each node. Primary sorting is carried out based on the relation priorities. The indeclinables such as emphatic particles, and conjuncts are left out as their positions are fixed with respect to their parent node. If there are relations with equal priorities at any level, secondary sorting is done based on the word order (ID) in the original sentence.

Linearizing the Tree

The sorted dependency tree is linearized to get the anvaya order. The tree is traversed using post-order-iteration and each node is added to the linear order pattern.

10. Description of the any Product/Process developed during the project

A module for generating the Anvaya order

11. Outputs & Outcomes:

a. Technology Development

- **Technology developed at Lab scale/pilot scale/commercial scale:**
— Lab scale

- **Technology demonstration in field setup - Yes**

- **Technology Transfer to industry: If Yes (Name of the Industry)**
— Not Applicable

- **Buyers/End users of product:**

Any Sanskrit Scholar interested in understanding Sanskrit texts.

- **Affordability: The cost-estimate of the developed prototype**

Source Code Available under GPL for Free

Application available on the website for free

b. Knowledge Creation

- **Patent filings:** Nil

- **Participation in conference/ training workshops attended:**

The module is a part of the Online Sanskrit teaching programme with Samsaadhanii.

So far trained more than 500 users and around 500 Sanskrit teachers all over India.

c. Knowledge Dissemination

- **Describe the strategies used to disseminate the research findings and knowledge.**

— Adding a module in the Sanskrit teaching course.

— Making the source code available on GitHub

- **Examples of dissemination methods:**

- i. Conference presentations**

- ii. Workshops and training sessions**

- Conducted workshops and training sessions for the Teachers.

- This module was part of the course

- iii. Online resources (websites, open access repositories)**

- For open access of the modules:

- <https://sanskrit.uohyd.ac.in/scl>

- <https://scl.samsaadhanii.in/scl>

- For the source code: <https://github.com/samsaadhanii/scl/tree/master/MT/prog/>

anvaya

- iv. List of Courses developed**

- A module in teaching Sanskrit using Computational Tools available at

- https://sanskrit.uohyd.ac.in/IOE-2024/lecture_notes.html

- v. Social media/other media outreach** — NA

- ***d . Publications***

- Research Papers**

- Books**

- Articles**

- Poster presentations Journals published through IKS center /**

- Project Magazines**

Type of Publications	Number of Publications	List of Publications	Authors & Co-Authors	Theme of the publication	References/ links of the Publications (link not required for books)
----------------------	------------------------	----------------------	----------------------	--------------------------	---

Mr Sanal Vikram who was employed as a project fellow in this project is working on the journal paper. His PhD thesis is on generating Anvaya.

e. Describe How the project resulted in integration of IKS and Contemporary knowledge

This project provided us an opportunity to use the traditional knowledge in building a technology. The traditional wisdom of the word order in a prose helped us develop the algorithm suitable to transform a verse into a prose.

f. Knowledge Repository

- Describe the structure and organization of the knowledge repository
Tall the programmes are available at <https://github.com/samsaadhanii/scl/tree/master/MT/prog/anvaya>
- Explain how users/public can access and utilize the knowledge repository.

The users can directly access them through GitHub. The programe is written in Python scripting language. The input and output conventions for the module will be added soon.

For using the anvaya generator, the weburl is

<https://sanskrit.uohyd.ac.in/scl>

<https://scl.samsaadhaniin/scl>

g. Capacity building

- Officers/faculty

— NIL

- Temporary manpower recruited/trained : One

संस्कृत - हिन्दी अनुवादकम् अनुभाषकञ्च (Sanskrit-Hindi Accessor Cum Machine Translator)	
संसाधनी- Samsādhani Updated on 3rd Feb 2024	
संस्कृत वाक्यम्	हिन्दी-अनुवाद
1. (मूलम्) अयोध्या-नगरे हस्तेन धनं ददाति ब्राह्मणाय दशरथ-पुत्रः रामः .	1. (मूलम्) अयोध्या नगरात् हाताने धनाला देतो ब्राह्मणाला दशरथ पुत्र राम
1. (अन्वयः) अयोध्या नगरे दशरथ पुत्रः रामः ब्राह्मणाय हस्तेन धनं ददाति .	1. (अन्वयः) ^अयोध्या नगरात् ^दशरथ पुत्र राम ब्राह्मणाला हाताने धनाला देतो

h. Results & Analysis with figures

संस्कृत - हिन्दी अनुवादकम् अनुभाषकञ्च (Sanskrit-Hindi Accessor Cum Machine Translator)	
संसाधनी- Samsādhani Updated on 3rd Feb 2024	
संस्कृत वाक्यम्	हिन्दी-अनुवाद
1. (मूलम्) दृष्ट्वा तु पाण्डव-अनीकम् व्यूढम् दुर्योधनः तदा आचार्यम् उपसङ्गम्य राजा वचनम् अब्रवीत् .	1. (मूलम्) देख कर तो पाण्डव-सेना के फुलाया हुए को दुर्योधन ने तब आचार्य के पास जा कर राजा वचन को कहा -
1. (अन्वयः) तदा राजा दुर्योधनः पाण्डव अनीकम् व्यूढम् दृष्ट्वा तु आचार्यम् उपसङ्गम्य वचनम् अब्रवीत् .	1. (अन्वयः) तब राजा दुर्योधन ने पाण्डव सेना के फुलाया हुए को देख कर तो आचार्य के पास जा कर वचन को कहा -

- Present your findings in a clear and organized manner

संस्कृत - हिन्दी अनुवादकम् अनुभाषकञ्च (Sanskrit-Hindi Accessor Cum Machine Translator)	
संसाधनी- Samsādhani Updated on 3rd Feb 2024	
संस्कृत वाक्यम्	हिन्दी-अनुवाद
1. (मूलम्) दृष्ट्वा तु पाण्डव-अनीकम् व्यूढम् दुर्योधनः तदा आचार्यम् उपसङ्गम्य राजा वचनम् अब्रवीत् .	1. (मूलम्) पाहून तर पाण्डव सेना फुलवलेला दुर्योधन तेव्हा गुरुला थांबून राजा वचनाला बोलला
1. (अन्वयः) तदा राजा दुर्योधनः पाण्डव अनीकम् व्यूढम् दृष्ट्वा तु आचार्यम् उपसङ्गम्य वचनम् अब्रवीत् .	1. (अन्वयः) ^तेव्हा ^राजा दुर्योधन ^पाण्डव ^सेना ^फुलवलेला पाहून तर गुरुला थांबून वचनाला बोलला

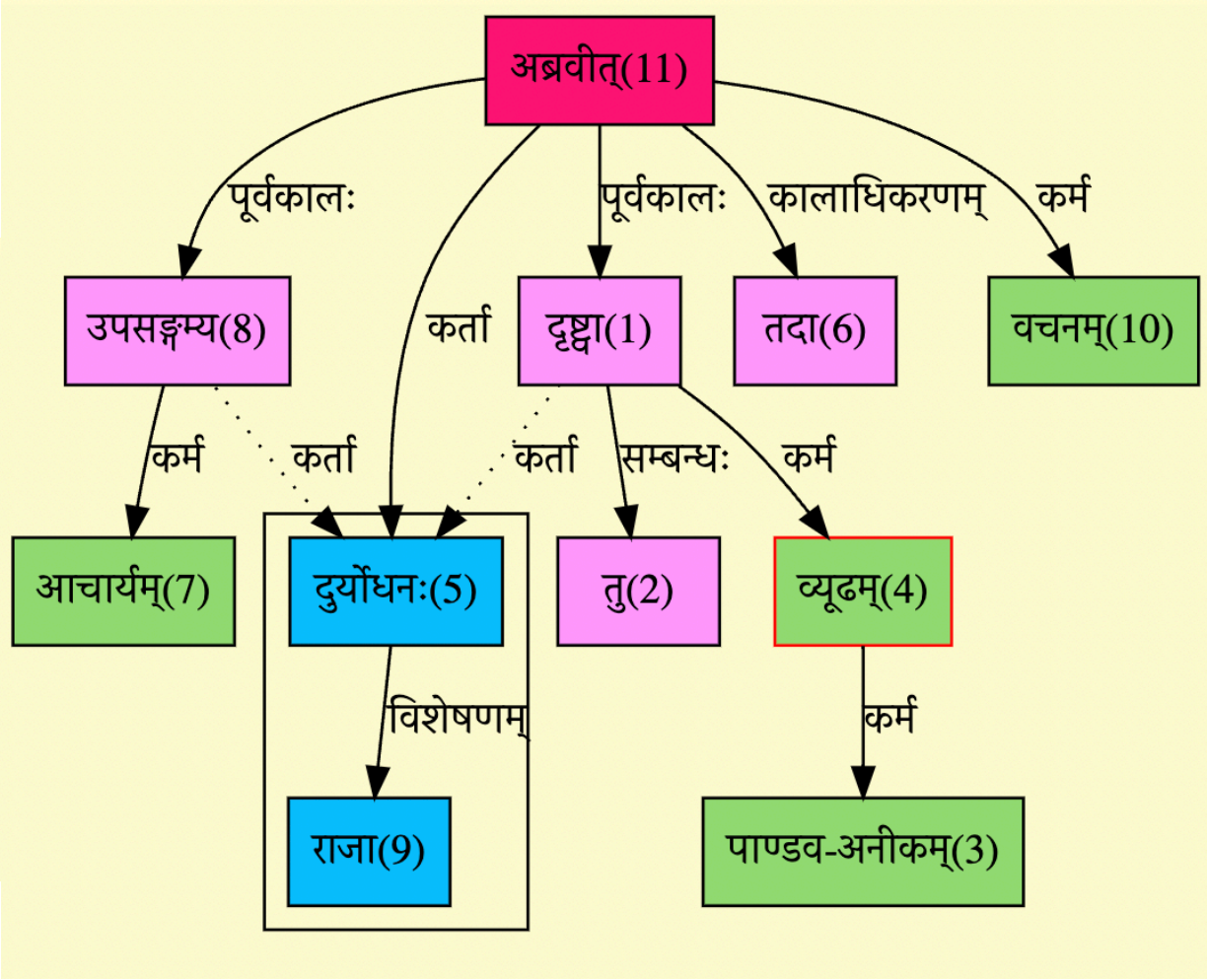
Given a Sanskrit verse, the programme generates an anvaya (also known as Dandaanvaya) or the prose form - both in original Sanskrit as well as in the translations in Hindi/Marathi/Telugu.

Below are a few screen-shots of the anvaya of Sanskrit and that of the translation in Hindi/Marathi.

संस्कृत - हिन्दी अनुवादकम् अनुभाषकञ्च (Sanskrit-Hindi Accessor Cum Machine Translator)	
संसाधनी- Samsādhani Updated on 3rd Feb 2024	
संस्कृत वाक्यम्	हिन्दी-अनुवाद
1. (मूलम्) अयोध्या-नगरे हस्तेन धनं ददाति ब्राह्मणाय दशरथ-पुत्रः रामः .	1. (मूलम्) अयोध्या-नगर में हाथ से धन देता है ब्राह्मण को दशरथ- पुत्र राम -
1. (अन्वयः) अयोध्या नगरे दशरथ पुत्रः रामः ब्राह्मणाय हस्तेन धनं ददाति .	1. (अन्वयः) अयोध्या नगर में दशरथ पुत्र राम ब्राह्मण को हाथ से धन देता है -

The translation into Hindi for the word व्यूढं has gone wrong, since the work on Word Sense Disambiguation is not yet started. Once the work on the Sanskrit-Marathi/Hindi Word Sense Disambiguation proceeds, the quality of the output will improve.

However, the word order in both Marathi and Hindi looks correct, which is the output of our work.



The change in word order is essentially obtained through the post-order traversal of the tree, by attaching appropriate weights to the relations on the sibling nodes. Below are the trees of these sentences.

j. Achievements/ highlights of the project

An integration of the Anvaya generation module has improved the quality of the translation.

j. References

- V. S. Apte. 1925. *The Student's Guide to Sanskrit Composition*, 9 edition. The Standard Publishing Company, Girgaon, Bombay.
- R. N. Aralikatti. 1991. A note on word order in modern spoken Sanskrit and some positive constraints. In Hans Henrich Hock, editor, *Studies in Sanskrit Syntax*, pages 13–18. Motilal Banarsidass Publishers.
- Brendan S. Gillon. 1996. Word order in Classical Sanskrit. *Indian Linguistics*, 57(1):1–35.

Brendan S. Gillon. 2005. Subject predicate order in Classical Sanskrit. In Philip Scott, Claudia Casa- dio, and Robert Seely, editors, Language and Gram- mar: Studies in Mathematical Linguistics and Natu- ral Language, pages 211–225. Center for the Study of Language and Information.

Amba Kulkarni. 2019. Sanskrit Parsing Based on the Theories of Śābdabodha. D K Printworld.

Peter Scharf, Anuja Ajotikar, Sampada Savardekar, and Pawan Goyal. 2015. Dis- tinctive features of poetic syntax preliminary results. In Sanskrit Syntax, pages 305—324. Sanskrit Library, USA.

J. Frits Staal. 1967. Word Order in Sanskrit and Uni- versal Grammar, volume 5 of Foundations of Lan- guage Supplementary Series. D. Reidel Publishing Company, Dordrecht-Holland.

Gary A. Tubb and Emery R. Boose. 2007. Scholastic Sanskrit: A Handbook for students. The American Institute of Buddhist Studies at Columbia University in the City of New York, New York.

k. Appendices

- NIL

l. Project Outcomes Contribution to (Bullet points. No more than 250 words in each category)

- **Dr̥ṣṭi:**

The Indian theories of Verbal Cognition, the tradition of commentaries, is relevant for Language Technologies

- **Jñāna:**

The knowledge existing in the shaastras need to be polished and adapted for the technology. This task is not trivial. Rather it is as hard as mining gold.

- **Laukika-prayojana:**

There is a dire need to provide an access to the trasure available in Sanskrit shastras. The use of scientific knowledge in the Indian linguistic sciences provides a scientific base for building the technology so that the knowldge can be accessed by a common man.

Amber?

Amba Kulkarni

Pavankumar Satuluri

Name of PI

Name of Co-PI

Head of the Institution

Signature of PI

Signature of Co-PI

Signature