Saṃsādhanī

Frequently Asked Questions

Team Saṁsādhanī

April 5, 2021

Contents

1 General

1.1 What is Saṃsādhanī?

Saṁsādhanī is a platform for Sanskrit Computational Linguistics developed at teh University of Hyderabad, under the guidance of Amba Kulkarni. It reflects the research undertaken at the Department of Sanskrit Studies at University of Hyderabad. Saṃsādhanī provides various tools for analysis of Sanskrit texts. These tools are

- Basic Tools:
  - **Morphological Analyser** (To get the word analysis)
  - **Morphological Generator** (To see various vibhkti forms of a prātipadika, to see various verbal forms in different lakāra, puruṣa, vacana etc. to see new prātipadikas derived from a given prātipadika/dhātu)
  - **Sandhi Joiner** joins any two Sanskrit *words* and shows the Paninian sutras involved
  - **Sandhi Splitter** provides the possible splittings of a compound word or joint words.
  - **E-readers**
    - Saṅkṣepa-rāmāyaṇam
    - Srimad-bhagavadgītā
    - Śiśupālavadham (only available on http://sanskrit.uohyd.ac.in/scl)
  - **Lipi-parivartanikā** - A simple tool for transliteration allows you to transliterate Sanskrit texts among various schemes.
- Advanced Tools:
  - **Anusaaraka (Sanskrit-Hindi MT)** provides a tool to assist a reader in understanding a Sanskrit text with the help of Hindi glosses, following the steps of śabda-viśleṣaṇa and ākāṅkṣā.
  - **Samasta-pada-vyutpādikā** generates a compound word from the prātipadikas following the Paninian process.

- **Nyāyacitradīpikā** is a tool to understand the Navya Nyāya expressions. The tool segments a NN expression, semi-automatically parses the string and renders the expression using a graphical interface.
- **Amarakośajāla** - This is an e-version of the most referred Sanskrit thesaurus Amarakośa, not only allows the search of synsets but also provides access to various other related words in the Amarakośa.
- **Concordance of Pāṇinian Dhātuvṛttis** displays the dhātus with pada and seṭ/aniṭ information with links to various vṛttis.
- **Aṣṭādhyāyi Simulator** shows the derivation of nominal forms of a noun by simulating the process Panini has employed in his Aṣṭādhyāyi.

1.2 What are the website URLs for the Saṃsādhanī Platform?

- http://sanskrit.uohyd.ac.in/scl – hosted at the University of Hyderabad
- http://scl.samsaadhanii.in/scl – hosted on the cloud
- http://sanskrit.inria.fr/scl – hosted at INRIA Lab in Paris, under the collaboration project between the University of Hyderabad and INRIA, Paris

In addition, there is an independent effort by Prof Gŕard Huet at Inria, Paris. The Sanskrit Computational tools developed by him are available at http://sanskrit.inria.fr.

1.3 What is Heritage Platform?

Heritage Platform is a platform developed by Prof Gérard Huet, Emeritus Prof at Inria, France. This platform provides various computational tools such as morphological analyser, morphological generator and sandhi joiner and splitter. In addition it has two dictionaries - Sanskrit-French(Heritage) and Sanskrit-Hindi (Monier-Williams).

1.4 Communication between Heritage and Saṃsādhanī

The Heritage Platform has the best segmenter for Sanskrit, while Saṁsādhanī is the only platform for Sanskrit that has the sentence analyser. The collaboration between the two researchers Gérard Huet and Amba Kulakrni has resulted into a computational system that accepts any Sanskrit sentence and produces the sentenctial analysis. Thus the two systems together help a reader to understand the Sanskrit texts.

2 Understanding Words

2.1 When do I use a Morphological Generator?

If you know the prātipadikam / dhātu and would like to know various forms of it, then you use the morphological generator. For example, if you would like to know what is the Tṛtīyā vibhakti ekavacana form of the prātipadikam **वचस्**, then you may use the Noun form generator. It generates the declensions (all forms in different vibhaktis). Enter this prātipadikam and choose appropriate **लिङ्गम्** . In response, the machine will produce a table showing all the forms of the word **वचस्**. To see the conjugation (different forms of a dhātu in different lakāras, puruṣa and vacanam), we use the Verb generator.

2.2 When do I use a Morphological Analyser?

If you do not know the analysis of a word - what is the prātipadikam, what is the vibhakti, or which lakāra is there, which vacanam, which puruṣa, etc., then you use the morphological analyser.

2.3 Where can I get the Sanskrit equivalents of morph analysis produced by the Heritage Engine?

The mappings between the Sanskrit Heritage (SH) and Saṃsādhanī (UoH) notation scheme is available at https://sanskritlibrary.org/helpmorphids.html

3 Segmenter

3.1 Is using the segmenter mandatory, when one can split the śloka themselves and use Anusārakam directly?

No. Use of segmenter is not mandatory. Segmenter is only meant to help us when we are unable to find the word boundaries by ourselves.

3.2 How do I split the words with the help of a machine?

There are two segmenters available for splitting the words.

1. Sandhi splitter in **संसाधनी** website.

   This splitter works well with a group of 3-4 words. However, it yields better results only if you specify whether the words thus joined is a compound or not. Further it does not handle the visarga sandhi that involves a space in between.

2. Segmenter in Heritage Platform.

   This is the most efficient segmenter. It can work well with a text involving several words joined together, involving visarga sandhi, etc.

   In order to use this effectively, you need some training regarding the various colours used to refer to various lexical categories. Refer page 7.

दृष्ट्वा तु पाण्डव-अनीकम् व्यूढम् दुर्योधनः तदा ।
आचार्यम् उपसङ्गम्य राजा वचनम् अब्रवीत् ॥

So we discourage the use of vertical bars in such cases where the complete śloka represents a single meaning.

However, if you are using Heritage Platform for segmentation, in the above example, the vertical bars are necessary. Because, the first line ends with **आ** and the next line starts with **आ**. Thus machine detects a potential sandhi formation, and signals an error in such cases. But if you use the vertical bar, then machine does not signal any error message.

3.3 Do we have to split all the components in a compound word?

Not necessarily. If you do not know all the components, and hence even if you split only the last component, the Anusārakam will work and produce the analysis.

For example, the compound **तपस्स्वाध्यायनिरतम्** can be split simply into two components as **तपस्स्वाध्याय-निरतम्**. You need not split it as **तपस्-स्वाध्याय-निरतम्**

3.4 How to save the segmentation?

Once you are done with the segmentation, click on UoH Analysis mode. Now here you may copy the segmentation and save it in a note pad or as a word document for further use, by pasting it there. You will note that the words, when pasted are sepatrated by a Tab. You mat remove the tabs and add blank spaces in between. Also after every 'yellow' colored word, add a hyphen.

4 Color Coding

4.1 What is the relevance of the colour-codes in Heritage and Saṁsādhanī?

The colour code gives you visual information about the lexical category (noun/verb/avyaya etc.) of the word. So even without looking at the word analysis, simply by looking at the

colour code, it is possible for you to choose the correct segmentation/correct analysis in the given context, manually.

4.2 Where do I get the information about the colours used in Saṃsādhanī?

The following colour information is available at http://scl.samsaadhanii.in/scl/SHMT/anu_help.html

Nouns in prathama vibhakti	deepskyblue
Nouns in dvitiiyaa vibhakti	GreenYellow
Nouns in tritiiyaa vibhakti	turquoise
Nouns in caturthii vibhakti	light skyblue1
Nouns in pancami vibhakti	lightgreen
Nouns in shashthii vibhakti	skyblue
Nouns in saptamii vibhakti	slategray
Nouns in sambodhana	Aquamarine(GH)
Indeclinable	Mauve(GH)
Kriyapada	Cadmiumredlight

Figure 1: colour Code in Saṃsādhanī

4.3 Where do I get the information about the colours used in Heritage?

You may refer to https://sanskrit.inria.fr/manual.html. Below are the most frequently used colours and the categories associated with them.

- substantive/adjective forms (blue)
- vocative forms (green)
- finite verbal forms (red)
- indeclinable forms such as adverbs, conjunctions, prepositions (mauve)
- pronominal forms (light blue)
- initial part of compounds (yellow)

4.4 If I get a gray coloured word in the Heritage segmenter, how do I proceed?

The gray colour indicates that the word is not recognised by the machine. In such cases, you may click on the little red heart below it and simplify the word by removing some components, which you are familiar with, and submit it again.

4.5 When I submit a sentence/śloka to the Anusārakam, I get the analysis. But some words are displayed in white. What should I do in such cases?

If a word is in white colour, then it means that the machine has failed to analyse this word. In such cases

1. If you are confident that the word is meaningful, and grammatically correct, and is not a compound or not a combination of two or more words, report to us. This would enable us to improve our morphological analyser.

   In such cases, for getting a temporary solution, if this word is a **विशेषणम्**, you may drop it and get the analysis. If it is a verb, replace it with some other verb of similar expectancies (that is, replace a sakarmaka verb with sakaramaka, akarmaka with akarmaka, etc.), and get the graph. Then in the Excel sheet you may replace the substituted verb with the original verb, or insert the dropped word. When you insert the dropped verb, then the word numbers will change accordingly. You will have to change them manually.

2. If you think the word in white colour is a compound word, split the components with '-' in between, and try the sentence again.

3. If you think the word in white colour is a sandhi between one or more words, then split the words, this time with a space in between, and try the sentence again.

5 Sentence

5.1 How do I know how many sentences are there in a śloka?

Here are some examples.

- Example 1:
  Sanskrit śloka:

तस्य सञ्जनयन् हर्षम् कुरु-वृद्धः पितामहः।
सिंह-नादम् विनद्य उच्चैः शङ्खम् दध्मौ प्रतापवान् (1.12)

Hindi Translation: दुर्योधन के हृदय में हर्ष उत्पन्न करते हुए कुरुवृद्ध प्रभावशाली पितामह भीष्म ने सिंह के समान गरज कर जोर से शंख बजाया।

English Translation: His glorious grandsire (Bhishma), the oldest of the Kauravas, in order to cheer Duryodhana, now roared like a lion, and blew his conch.

In this śloka there is only one sentence, since there is only one finite verb **दध्मौ** (**बजाया** , blew)

- Example 2:
  गाण्डीवम् स्रंसते हस्तात् त्वक् च एव परिदह्यते।
  न च शक्नोमि अवस्थातुम् भ्रमति इव च मे मनः

  In this śloka there are 4 small sentences viz.

  1. गाण्डीवम् स्रंसते हस्तात्
  2. त्वक् च एव परिदह्यते
  3. न च शक्नोमि अवस्थातुम्
  4. भ्रमति इव च मे मनः

  Here there were 4 verbs (tinanta), and thus four sentences.

- Example 3:
  In the following śloka, though there are two verbs - **गृह्णाति** and **संयाति**, still this is a single sentence, because in the first sentence there is a word **यथा**, which has an expectancy for **तथा**.

  वासांसि जीर्णानि यथा विहाय
  नवानि गृह्णाति नरोपराणि
  तथा शरीराणि विहाय जीर्णानि
  अन्यानि संयाति नवानि देही

5.2 When to do an adhyāhāra?

When there is a necessity of a **क्रियापद**, we do its **अध्याहार**.

- Example 1:
  सः अध्यापकः
  $\Rightarrow$ सः अध्यापकः अस्ति

- Example 2:
  न काङ्क्षे विजयम् कृष्ण न च राज्यम् सुखानि च।

⇒ (Here we need to repeat the verb and negation as below, and split it into three sentences.)
न काङ्क्षे विजयम् कृष्ण
न च राज्यम् काङ्क्षे
न च सुखानि काङ्क्षे

- Example 3:
  अस्माकम् तु विशिष्टाः ये तान् निबोध द्विज-उत्तम
  ⇒ (Here we need to supply the verb **सन्ति** as below.)
  अस्माकम् तु विशिष्टाः ये सन्ति तान् निबोध द्विज-उत्तम

5.3 Can I enter the vertical bars daṇḍa (। , ॥) ?

Yes, you can enter the vertical bars. Remember that these are different from | and || . Also note that ।। is different from ॥.

In the **संसाधनी**, these will be treated as sentence boundaries. So in the following **श्लोक**, machine will consider the two lines as two different sentences, and will fail to produce any analysis for the first part since it does not have any verb in it.

5.4 Do we have to keep all the viśeṣaṇas (adjectives) before the viśeṣya ?

If you want to run the Anusārakam in a prose mode, then it is necessary to bring all the **विशेषण** (adjectives) before the **विशेष्य**. But if you run it is śloka mode, then this order does not matter.

6 Sentence Analysis

6.1 What is a kāraka relation?

A relation between the activity and its participant is called a kāraka. There are 6 types of kāraka relations. They are kartā, karma, karaṇa, sampradāna, apādāna, and adhikaraṇa.

For example, in the sentence **रामः नगरे कोशात् हस्तेन ब्राह्मणाय धनं ददाति**, the relations of various participants with the activity are as follows.

Here the numbers denote the word indices in the sentence.

Remember that there can not be more than one participant having the same relation. If there are two kartās, then one of them should be a **विशेषणम्** for the other, or both of them are kartā, together, joined by a **समुच्चयद्योतक** such as **च / तथा / अपि**. Same thing holds good for **करण, सम्प्रदान** and **अपादान** as well.

1	रामः	कर्ता,7
2	नगरे	अधिकरणम्,7
3	कोशात्	अपादानम्,7
4	हस्तेन	करणम्,7
5	ब्राह्मणाय	सम्प्रदानम्,7
6	धनं	कर्म,7
7	ददाति	

However, there can be more than one **अधिकरण**, as in
रामः भारत-देशे अयोध्या-नगरे सायंकाले पञ्च-वादने पाठशालात् गृहं गच्छति.

Here there are 4 **अधिकरण**s - **भारत-देशे अयोध्या-नगरे सायंकाले पञ्च-वादने** . The first two of them indicate the space and the last two, time. Hence they may be termed as **देश** and **काल अधिकरण** respectively.

Similarly there are a group of verbs which may have 2 karmas. For example, in the case of **ब्रू**, answers to 'whom' and 'what' would result in two karmas. These are called **गौण-कर्म** and **मुख्य-कर्म**.

6.2 What is a kāraketara relation?

A kāraketara relation is between the objects that do not participate directly in the activity. For example, the 'purpose', 'reason', kinship relation, part-and-whole relation as in **वृक्षस्य शाखा**, or posessor-possessee relation as in **रामस्य पुस्तकम्**, the relation between two activities - such as whether one precedes the other, or if they are simultaneous etc., the relation of the verb with the addressee, the adjectival relation, the adverbial relation, are all kāraketara relations.

6.3 Where can I get the list of all relations that are used in the Saṁsādhanī?

The list is provided in the file named `relations_list.pdf` in `Lecture 13` folder; the Google drive link is given below—
https://drive.google.com/drive/folders/1x7JQkp9szJhZv7Uv7hPVce-2-JM_hMG2.

This folder also contains English and Sanskrit guidelines providing examples of various relations.

6.4 If Saṁsādhanī fails to analyse a word, how to proceed?

Try the sentence / śloka on the Saṁsādhanī - Anusārakam page with the Heritage splitter or directly try it on the Heritage website with the Reader.

6.5 All the words in my śloka are analysed by the machine. But still I do not get any graph. How to proceed?

In such cases, try to remove the adjectives, and check if you can get the analysis. Still if you do not get any analysis, then save the annotation. The kāraka analysis column would be empty. Enter the relations manually.

6.6 Why I am not getting the tree for a given sentence/verse?

There are mainly two reasons for not getting the tree, even though machine has analysed all the words. These reasons are

- Machine has failed to establish any relations for one or more words. This typically is due to incompleteness of the rules / data.
- Machine fails to get a tree from among all the relations (edges) it has established between various words (nodes). This typically is due to the free word order. If you manually change the word order to a Daṇḍānvaya, you get the analysis.

6.7 Is it possible to generate an anvaya through the machine?

The software that will generate an anvaya is still in research mode. It will take some time to make it available on the Saṁsādhanī platform.

7 Submission

7.1 Once the analysis is done and we get a Tree, what is the next step?

Cheers! You have reached the last step. Now you click on the 'Save Annotation' link at the bottom of the page. This will result into one of the following two scenario, as per the settings on your machine.

- You will be prompted to save the annotation in an Excel sheet.
- This will open the file as an excel sheet.

You may save this sheet with the following naming convention.

If your group id is G007, your sarga is 8, your śloka is 108, and this śloka consists of 3 sentences, and this is the analysis of the second sentence, then the name of the xlsx file would be `7_8_108_2.xlsx`

7.2 I have used the anvaya for analysis. Can I submit the sheet produced with the anvaya?

Yes. But you should also add the śloka word order under the column 'Poem'.

7.3 Should I submit the Khaṇḍānvaya as well?

This is not expected.

7.4 If I have to add some extra words such as 'asti' etc. then how should I number them, when I convert it to a śloka order?

Add the 'asti' etc. immediately before the next sentence/clause begins, and assign the numbers in normal order. Put the words that you have inserted manually in ().

For example, I insert 'santi' after the word 'ye' in the following BhG śloka.
अस्माकम् तु विशिष्टाः ये तान् निबोध द्विज-उत्तम resulting into
अस्माकम् तु विशिष्टाः ये सन्ति तान् निबोध द्विज-उत्तम
And the word सन्ति will have the index 5.

Now after you get the Excel sheet, put () around the words सन्ति, to indicate that this word is inserted by you, and is not the part of original śloka.

7.5 In the final Excel sheet I submit, should I remove the inserted words?

No. Do not remove the inserted words. Put these words in Parenthesis - i.e. round brackets viz. (), after getting the Exce sheet.

7.6 If I run the śloka after removing certain words that are not recognised, then how should I submit the final Excel sheet?

You should add the removed words manually at appropriate place in the Excel sheet, and also change the word indices accordingly.

7.7 Please tell us what should be done in the case where we change the word order to get the chart? How to retain the order of words as per the śloka?

If you have changed the word order then in the excel sheet, this word order would be reflected. Now in the column with heading 'Poem', put the word order according to the śloka.

7.8 How do I save the tree diagram?

Rightclick on the image and then there will be an option to save it.

7.9 How to do the anvaya?

Usually, the anvaya would be available in the commentaries. You may use that anvaya. However, make sure that this anvaya does not have unnecessary extra words. Adhāhāra of certain words such as 'asti' etc are fine. Also note that the viśeṣaṇas and viśeṣyas are not intervened by other words that are ot related to them. The main verb is at the end. Viśeṣaṇas are before the viśeṣyas.

7.10 If I add some extra words as an adhyāhāra, how to mark them in the Excel sheet?

After getting the Excel sheet, put these extra words in parenthesis, manually. Also when you provide the word numbers in the śloka form, assign numbers to these extra words as well.

7.11 Should I add the word numbers in the Poem column manually?

You add the word numbers only if the excel sheet corresponds to the anvaya and not the śloka.

7.12 Should I correct the Hindi meanings?

It is optional and left to you.

7.13 Can I add English meanings?

This is also optional.

8 Error

8.1 Why is "Internal Server Error" message displayed while using the Heritage splitter?

This message is displayed when the handshake between the Heritage splitter and the Saṁsādhanī is not smooth. This may be due to insufficient available memory, or it may be due to the failure of Saṁsādhanī to understand the segmented data produced by Heritage segmenter. In either case, you try the same example once again after some time, on the same server, or on another server. If the problem persists, report to ambapradeep@gmail.com providing the śloka/sentence.

8.2 The machine is very slow and takes long time to provide the anaysis. Why is it so?

There are two reasons for the 'slowness'. The first one is the huge number of hits (typically 50-100 ślokas at a time). The second reason is the huge number of solutions due to the śloka order.

If you give the anvaya, the number of possible answers reduce, and machine can provide you the solution quickly. However, with the śloka order, since the number of possiblities explode (since machine does not have much knowledge of the yogyatā), it takes a log time to produce the solutions. Typically the number of solutions in śloka order are 50-100 times more than the possible solutions in Anvaya order.