Reaching the Unreached

Through

Science & Technology

(Collection of Articles)

Accrtedited by NAAC
with "A" Grade

Accrtedited by IAO

Certified by ISO

Moulding Character & Careers

e Indian Science Congress Association
Bengaluru Chapter

ence Congress Association
Bengaluru Chapter

dian Science Congress Ass

SURANA COLLEGE
PERMANENTLY AFFILIATED TO BANGALORE UNIVERSITY

GDA FOUNDATION

SURANA EDUCATIONAL INSTITUTIONS

No: 16, South End Road, Bengaluru-560004
Phone: 080-26642292 / 26541095
Web: www.suranacollege.edu.in

Surana Vidyalaya, Chandapura Campus

Surana Independent PU College, South End Campus

Surana Independent PU College, Kengeri Campus

Surana Independent PU College, Peenya Campus

Surana College, South End Campus

Surana College, Kengeri Campus

Surana College, Peenya Campus

Surana Evening College, South End Campus

Surana Evening College, Kengeri Campus

Moulding Character & Careers

TRACK - 6:
Languages: Sanskrit

TRACK - 6:
Languages: Hindi

Morphological Generator for PĀLI

Swati Basapur
Research Scholar, Shastra Faculty
Karnataka Samskrita University
dswatisrini@gmail.com

Shivani V
Associate Professor, Shastra Faculty
Karnataka Samskrita University
shivani.ksu@gmail.com

Shrinivasa Varakhedi
Professor and Dean, Shastra Faculty
Karnataka Samskrita University
shrivara@gmail.com

Intrtoduction

Pāli is a language of Buddha's Upadeśa or teachings collected in Tipiṭakas namely Suttapiṭaka, Vinayapiṭaka and Abhidhammapiṭaka. There are bhāṣyas or commentaries written on these Tripiṭakas. Pāli has a rich heritage of literary works which need to be explored. Pāli computer linguistics efforts are required considering growing interest in Buddha's teaching in India and abroad. Pāli being highly inflectional language, generator for this language will help create more learning and teaching tools for Pāli and allied languages. This paper is about the morphological generator for Pāli.

Pāli

Pāli is the name given to the language of the texts of Theravāda Buddhism, although the commentarial tradition of the Theravādins state that the language of the canon is Māgadhī, the language supposedly spoken by the Buddha Gautama. For the linguistics, Historical research, Buddhism, Buddhist philosophy and such major factors are the reasons why Pāli should be studied and paid attention to. Therefor it is essential to have a modern system to study Pāli . Buddha taught his principles through Pāli which was spoken language then. Now we do not have many people who understand Pāli. This system is a first step towards bringing it closer to people in modern era.

Morphological Generator

In linguistics Morphology is study of words, their formation and relationships with other words.In grammar, inflection is the modification of a word to express different grammatical categories such as tense, case, voice, aspect, person, number, gender, and mood. The inflection of verbs is also called conjugation, and inflection of nouns, adjectives, adverbs, pronouns, determiners, participles, prepositions, postpositions, numerals, articles etc., as declension. Languages can be highly inflected (such as Latin, Greek, Spanish, Biblical Hebrew, and Sanskrit), or weakly inflected (such as English). Pāli also is highly inflectional language and suitable for morphological generation and analysis. Morphological generation means given the morphological information of a word base in terms of number, category, stem, and so on, the intended word form is retrieved. For example:

Input

      pratipadika: deva
      Category : nāmapada

Gender : Pumliṅga
Vibhakti : Tṛtīyā (instrumental case)
Number : ekavacana
Output of generator : देवेन

The other type of generation is given a word base, generating all its possible morphological forms.

Previous Work:

No complete Natural Language Processing Systems are available for Pāli. The ones available are: Pāli lookup System by Aukana Trust We could not find downloadable version. But it is not available for public. Another one is Pāli morphological analyzer and generator developed by David Alfter. It is not downloadable. Hence could not evaluate both the systems.

Morphological generator for Pāli :

Grammar books on Pāli language always discuss Pāli grammar with Sanskrit background. It was our Sanskrit background that prompted us to take up this task of developing Morphological generator for Pāli. We took it as advantage to learn Pāli grammar and explain Pāli morphologically. The strong relationship between languages also became a basis for us develop this tool based on Sanskrit Morphological analyzer and generator. Though Pāli is similar to Sanskrit in many ways, it has its own peculiarities. Therefore it needs its own computational tools. Pāli is written in various scripts all over the world. This tool takes Unicode input. Following are the constituent blocks of this tool.

- **Nominal Declension Generator**

Pāli is highly inflectional language as Sanskrit. Pāli does not have dual form. It has only singular and plural forms .This means nominal base (prātipadika) + suffix = nominal form.

E.g. बुद्ध + सि = बुद्ध + ओ = बुद्धो

Following is a table of nominal suffixes.

		एकवचन	बहुवचन
पठमा		सि	यो
आलपना		अ,आ	आ
दुतिया		अं	यो
ततिया		ना	हि
चतुत्थि		स	नं
पञ्चमी		स्मा	हि
छट्ठी		स	नं
सत्तमी		स्मिं	सु

After noticing this it was decided to use paradigm table as is the case with Sanskrit , because all akārānta pulliṅga i.e. ending with vowel 'a' behaved the same. So for akārānta pulliṅga we took 'deva' as standard and this was the case with most of the grammar we referred to.So the paradigm table for akārānta pulliṅga words looks like this:

	देवपु	
विभक्तिः	एकवचन	बहुवचन
पठमा	देवो	देवा
दुतिया	देवं	देवे
ततीया	देवेन	देवेहि/देवेभि
चतुत्थी	देवस्स/देवाय	देवानं
पञ्चमी	देवा/देवस्मा/देवम्हा/देवतो	देवेहि/देवेभि
छट्ठी	देवस्स	देवानं
सत्तमी	देवे/देवस्मिं/देवम्हि	देवेसु
आलपन	देवदेवा	देवा

During paradigm table creation following were the observations :All nominal roots are ending in vowels.

Akarant nominal roots	देव , अत्त , राज , पुम , युव
Ukarant nominal roots	भिक्खु , दातु , सत्थु
Akarant Napumsakalinga	मनफल

Though in above examples each row nominal roots have same ending but paradigm tables are different.e.g. ततीयाविभक्ति singular forms of some pratipadikas are given below.

देव	राज	अत्त	पुम	युव
देवेन	रञ्ञा,राजेन,राजिना	अत्तेन,अत्तना	पुमाना,पुमेन,पुमुना	युवाना,युवानेन,युवेन

We can see nakārānta prātipadika influence in form रञ्ञा and akārānta prātipadika influence in form राजेन and conjunct ञ is simplified by insertion of vowel in the form राजिना.

We have created separate paradigm table for special words. And based on Sanskrit we are recognizing approximately the ending varnas for e.g. words declining like raja (rājan in Sanskrit). Similar to Sanskrit there is a ganapātha in Pāli. So with the help of ganapātha we have grouped some paradigms like मनादि gana has तप, तेज, पय ,रस, यस, ओज etc. Here these are all akārānta

napumsakalinga but are grouped under one gana so they are assigned paradigm 'मन'. But for दातृसत्थु do not come under any gana where we have to identify which ones behave like दातृ and which ones like पितृ assign them respective paradigm. After generation we are testing against validated forms from grammar books that are available.

For reference we have used Pāli-English dictionary by Rhys Davids and William Stede which is taken as the basis for the lexical database. This dictionary has few issues though it helped us with collection of almost 16000 words. This dictionary can be regarded fairly as the standard, since there are only a handful of dictionaries for Pāli. We have manually added this data and processed computationally.

- **Verbal conjugation Generator :**

Verbal forms in Pāli are generated based on Sanskrit verbal form generation. In Pāli also verbal forms are derived by adding verbal suffix to verbal root.

e.g. नम + अ + मि = नमामि

verbal root + vikarana + verbal suffix = verbal form

Verbal suffix table for vartamana kala(present tense)is as follows:

	एकवचन	बहुवचन
पठमपुरिस्स	ति	अन्ति
मज्झिमपुरिस्स	सि	थ
उत्तमपुरिस्स	म	मि

It is seen here that verbal forms do not have द्विवचन form . Verbal forms have only एकवचन and बहुवचन forms. Also in Sanskrit all roots are either Atmanepadin or Parasmaipadin or Ubhayapadin . In case of Pāli , most of the times they assume both atmanepadin and parasmaipadin forms. But there is no specific rule as to which root is parasmaipadin and which is atmanepadin. Therefore for each verbal root we generated both forms namely parasmepadi forms and Atmanepadi forms. This led to over generation problems. As per most of the books atmanepadin forms are less frequent.Verbal forms available in grammar books at hand only could be validated.

Ganas are groups under which dhatus in Pāli are classified. Clearly similarity between Sanskrit and Pāli is seen here. Having three schools of grammar , It is observed that are followed in different ganas in different grammars. For e.g. In Moggalana vyakarana dhatus are classified into following ganas . Bhuvadi, Rudhadi, Divadi, Tudadi, Jyadi   , Kiyadi, Tanadi, svadi, churadi . Where as in Kaccayana they are classified into 7 ganas. Saddaniti school of grammar follows kaccayana but it has created separate gana for गह .So that makes it 8.After talking to experts we decided to follow Moggalana Vyakarana.

Lakara or tenses in Pāli with respect to Moggalana are Vattamana, Bhavissati, Parisamapti, anajjatana, parokkha, hetuhetumdboot, vidhiling and Anudnya.

Sample verbal forms table

- **Parasmaipadi forms**

	एकवचन	बहुवचन
पठमपुरिस्स	नमति	नमन्ति
मज्झिमपुरिस्स	नमसि	नमथ
उत्तमपुरिस्स	नमाम	नमामि

- **Atmanepadi verbal form table**

	एकवचन	बहुवचन
पठमपुरिस्स	नमते	नमन्ते
मज्झिमपुरिस्स	नमसे	नमव्हे
उत्तमपुरिस्स	नमे	नमाम्हे

Verbal roots from Saddaniti and Moggallana Dhatupatha were consolidated for the purpose of verbal forms generation. Around 1600 verbal roots are consolidated and above mentioned lakara forms generated.

Challenges

There are multiple schools of grammar as we understand. Kacchayana Grammar is the traditional one and is divided into 8 chapters based on combination, Declension, syntax, Compounds, Nominal derivatives, Unnadi affixes. But total number of sutras is 687.Moggalayana's grammar which is written in imitation of Kachchayana and has rules, vruttis and examples. Saddaniti is in line with Kaccayana. There is variation in terminology like Vidhiling in Moggallana is called Sattami in Kachchayana, Anujnya in Moggalayana is called Panchami in Kachchayana grammar. But Panchami and Saptami are Vibhakti also in Pāli as in Sanskrit. After talking to experts it was decided to follow Moggalayana School of grammar. Also due to unavailability of Standard book of Noun Declensions and Verb forms all the forms generated couldn't be tested. Validation of only forms which are in the grammar books for Pāli could be done. The way we have Dhaatu Ratnakara and RoopChandrika for various noun and verb form as standard books we are still in search of similar books for Pāli forms. Also for nominal forms standard paradigm tables were created from Moggallana vyakarana book.

Outcome of the Project:

The Morphological Generator can take Pali stems or roots and generate the word forms. A User friendly web based GUI (Graphical User Interface) has been created for this tool. By clicking the options it opens the generator. Front page of the Pali Web page is given below:
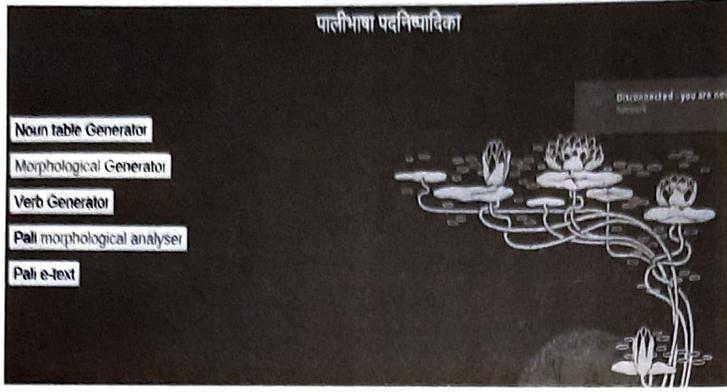
This is the Main Menu of the morphological generator tool. When morphological generator button is clicked following page is displayed where we can give morphological information.

Input:

Two options are there for giving input.

1. Giving morphological information such as gender, vibhakti, lakara, number etc. and getting that particular nominal/verbal form generated. Following is the webpage for input1.
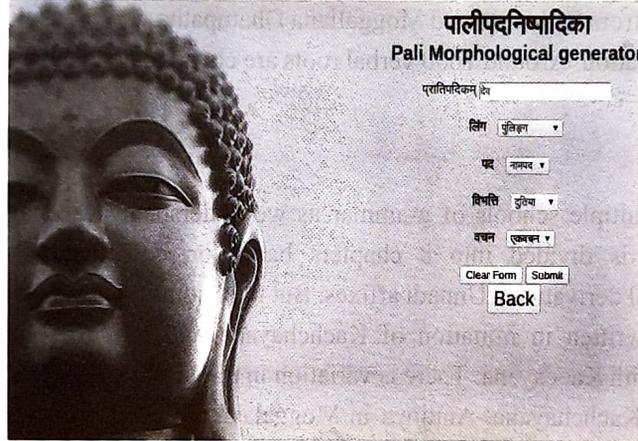
After clicking submit button output is displayed as follows.

देवं

प्रातिपदिकम् ==> देव

लिङ्गम् ==> पुंलिङ्ग

वचनम् ==> एकवचन

विभक्ति: ==> दुतिया

Giving nominal root and gender and getting a paradigm table generated

2. In the main menu when noun table generator button is clicked following page gets displayed.
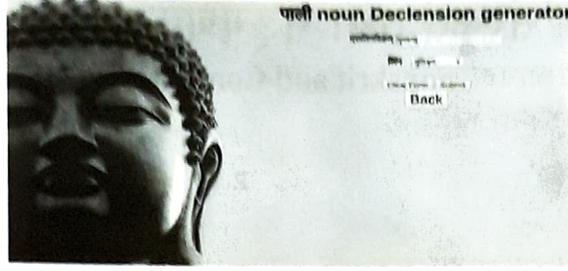
Generator output for गुणवन्तु

विभक्ति (विभक्ति)	एकवचनम्(एकवचनम्)	बहुवचनम्(बहुवचनम्)
पठमा (पठमा)	गुणवन्तु	गुणवन्तो गुणवन्तु
दुतिया (दुतिया)	गुणवन्तु	गुणवन्तो गुणवन्
तातिया (ततिया)	गुणवन्तुन	गुणवन्तुभि गुणवन्तुभि
चतुत्थि (चतुत्थी)	गुणवन्तुने गुणवन्तुरस	गुणवन्तां
पञ्चमी (पञ्चमी)	गुणवन्तुर्व गुणवन्तुस गुणवन्तुस्म	गुणवन्तुभि गुणवन्तुभि
छट्ठी (छट्ठी)	गुणवन्तो गुणवन्तस	गुणवन्तां
सत्तमी (सत्तमी)	गुणवन्तुभि गुणवन्तुस्मि	गुणवन्तुसु गुणवन्तुसु
आलपना (सम्बोधन)	गुणवन्तु	गुणवन्तो गुणवन्तो गुणवन्

This tool also generates all forms of given lakara of given verbal root.

It can also generate individual form if we give input of verbal root and morphological information of required form.

Future work to be done

We have been successful in creating generator for Pali with existing resources. Upasargas, samasas, taddhitas , unadi, derived dhatus are yet to be generated. This tool will be a stepping stone of analyzer and further a complete machine translation system.

Conclusion

Though Pali is similar to Sanskrit in many ways it still needs its own computational tools to lead way to future machine translation system for Pāli as well as allied prakrit languages. This study will create more awareness about these languages and more scholars so that information contained in these languages is not lost.

Acknowledgement:

Bibliography:

1. Childers, Robert Caesar (1875) "A dictionary of pāli language", Trubner and co., London 298 Davids, Rhys and William Stede (1921-25)
2. D'Alwis, James (1863) "Introduction to kachchayanas grammar of the Pāli language", Colombo
3. Dhammakitti, Bhikku (1996) "Pālivyākaraṇa : bālavatāra of Bhikku dhammakitti with hindi translation", 301 Bauddha bhārati, Varanasi.
4. "Pāli – English Dictionary", Pāli Text Society.
5. Kashyap, Jagadish Bhikshu (1940) "Pāli Mahavyākaraṇa", Mahabodhi Sabha, Saranath, Banaras.
6. Nyanatiloka, Ven. (1988) "Buddhist Dictionary:Manual Manual of Buddhist Terms and Doc- trines", Buddhist Publication Society, Kandy, Sri Lanka
7. Saddhatissa, (1949) "Saral Pāli Śikṣā", Mahabodhi society, Saranath, Banaras
8. Silananda, Venerable U "Pāli roots in Saddanīti", e-book.
9. Tungar, Na Va (1939) "Pāli Bhāṣā Praveśa", Samarth Bharat, Pune.