

Reaching the Unreached

Through

Science & Technology

(Collection of Articles)



Accredited by NAAC
with "A" Grade



Accredited by IAO



Certified by ISO

Moulding Character & Careers



GDA FOUNDATION

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

ISBN: 978-81-931907-1-5



80. A Study on the Effect of Cyber Interaction on Interpersonal Relationship of College Students in Bangalore City
Noor Nigar | 500
81. 'Working Plan' a valid tool to plan the Forest development and Forestry
Dr. Pampa Devi. B.; Harshita U Urs | 508
82. Political Intolerance in West Bengal: The Crisis of Illiberality
Jacqueline Elisha Gomes | 510
83. Politico-Economic Model of Sino-Indo-Pak Arms Race
Dr. Shukla Vidya Prasad | 515

TRACK - 6: **Languages: Sanskrit**

84. Ayussamskritam : A Computational Tool To Study Ayurveda Texts
Shivaja S. Nair; Srinivas Varakhedi; Shivani V. | 526
85. Morphological Generator for PĀLI
Swati Basapur; Shivani V.; Shrinivasa Varakhedi | 531
86. Sanskrit and Computer
Vidwan Naveen Bhat | 538
87. Samskritashastrapaathane Tantrajanopayogitha
Dr. Mallikarjun B. S. | 542
88. Samskritasthavijnanasya Tantrajanadvara Sampreshanam
Dr. Tirumal | 545
89. Samskrita and science
Padmaja D. S.; Shivananda L.B. | 547
90. Samacharasankethika Madhyamena Samskritadhyayana Adhyapanaridhi
Dr. K Varalakshmi | 549
91. The Role of Sanskrit and ICT in Strengthening Each Other
Dr. Selvappa | 556

TRACK - 6: **Languages: Hindi**

92. Aadhunik sanchaar maadhyam Aur yuva varg ka Shahar aur gauv par prabhav
Asst. prof. Simerjeet Kler | 559
93. Bhishm Sahani ke Natak- Samkalin Pariprekshya Mein
Rekha P. Menon | 564
94. Hindi Kshikshan Aur Bhasha Ke Vikas Mein Suchana Prodhhyogiki ka yogdaan: Ek Samiksha
Dr. Ranjeet Kumar | 569
95. Hindi Sahitya Mein Naatya Kruti Aur Rangmanch Shaily par Takniki Yug Ka Prabhav
Dr. Usha Rani Rao | 577
96. Hindi Sahity Me Natya Kruti Evam Rangamanch Shaili Par Takaneeki Yug Ka Prabhaav
Prathyush Kumar P. | 581
97. Jansanshar Madhyam aur Yuva Varg
Dr. Malati Prakash | 586
98. Vigyan aur takniki ke shaitra mein Hindi Bhasha
Dr. Veer Mani Panday | 590
99. Hindi Basha Jagat Mein jan-Sanchar Madhyamo ki Bhumika
Dr. Shyam sunder | 593
100. Hindi Sahitya Mein Naatya Kruti Aur Rangmanch Shaily par Takniki Yug Ka Prabhav
Dr. Mukesh Kumar Mishra | 597

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

Introduction

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. Therefore 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 : Pumlīṅga
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 Altfer. 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		स्सि	सु

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

	देवपु	
विभक्तिः	एकवचन	बहुवचन
पठमा	देवो	देवा
दुतिया	देवं	देवे
ततीया	देवेन	देवेहि/देवेभि
चतुथी	देवस्स/देवाय	देवानं
पञ्चमी	देवा/देवस्सा/देवम्हा/देवतो	देवेहि/देवेभि
छट्टी	देवस्स	देवानं
सत्तमी	देवे/देवस्मि/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	भिक्षु , दातु , सत्थु
Akarant Napumsakalinga	मनफल

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

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

We can see nakārānta prātipadika influence in form रञ्जा 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 ganapāᅇha in Pāli. So with the help of ganapāᅇha 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 दातुसत्थु do not come under any gana where we have to identify which ones behave like दातु and which ones like पितु 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
मञ्झिमपुरिस्स	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, hetuhetumboot, 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

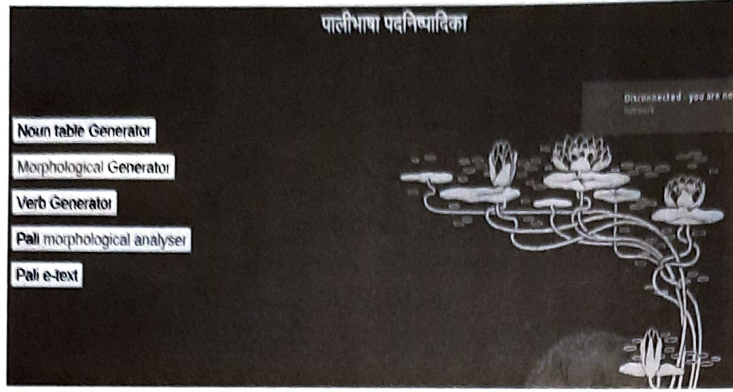
Verbal roots from Saddaniti and Moggallana Dhatupatha were consolidated for the purpose of verbal forms generation. Around 1600 verbal roots are consolidated and above mentioned lakara forms generated.

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:

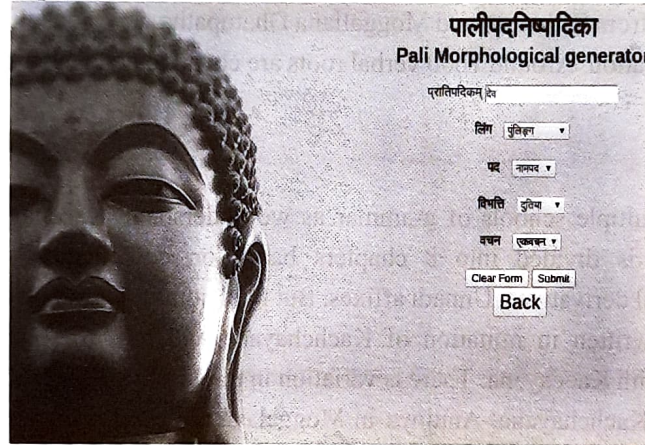


This is the Main Menu of the morphological generator tool. When morphological generator button is clicked following page is displayed where we can give morphological information.

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

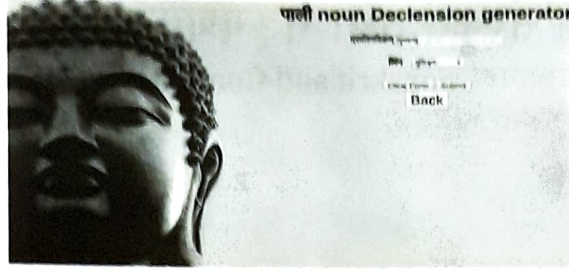
लिङ्गम् ==> पुल्लिङ्ग

वचनम् ==> एकवचन

विभक्तिः ==> 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)	लकार (लकार)	प्रकार (प्रकार)
एकवचन (एकवचन)	गुणवन्तु	गुणवन्तु
द्विवचन (द्विवचन)	गुणवन्तु	गुणवन्तु
तृतीया (तृतीया)	गुणवन्तु	गुणवन्तु
चतुर्थी (चतुर्थी)	गुणवन्तु	गुणवन्तु
पञ्चमी (पञ्चमी)	गुणवन्तु	गुणवन्तु
षष्ठी (षष्ठी)	गुणवन्तु	गुणवन्तु
सप्तमी (सप्तमी)	गुणवन्तु	गुणवन्तु
अष्टमि (अष्टमि)	गुणवन्तु	गुणवन्तु
नवमि (नवमि)	गुणवन्तु	गुणवन्तु
अष्टमि (अष्टमि)	गुणवन्तु	गुणवन्तु

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:

We thank Prof Amba Kulakarni for her invaluable guidance. We also acknowledge financial assistance by Pāli Institute Kalburgi, Karnataka for developing this tool.

Bibliography:

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

Pāli Sandhi – A Computational Approach

Swati Basapur

Research Scholar

Karnataka Samskrita University

dswatisrini@gmail.com

Shivani V

Professor

Karnataka Samskrita University

shivani.ksu@gmail.com

Sivaja S Nair

Post Doctoral Fellow

Karnataka Samskrita University

sivaja.s.nair@gmail.com

Abstract

For any Indian language, the accuracy of the morphological analyser, depends on the pre-edition of the input text. In Pāli language, like any other Indian language, the combination of words like sandhis and samāsas are frequently seen. This poses difficulty in the proper analysis of the source text. It is essential to have computational tools that help to split the words, useful in the analysis of the text. This paper discusses complexities involved in creating a computational grammar for sandhi tool in Pāli language.¹

1 Introduction

Pāli is a widely studied classical language, mainly because it is the language of Pāli canon. A growing interest in Pāli makes it important to develop computational tools for the language. Morphological analyser/generator is one such effort in this direction. All the combined words, (sandhis, samāsās, etc.) used in the text have to be manually split before using it as an input to the morphological analyzer in Pāli language. Since it is a tedious effort, pre-editing tools such as sandhi splitter/joiner and samāsa analyser were envisaged. Though similarities were observed in Pāli and Samskrita grammar, it was observed that Pāli grammar was much more complex. This paper discusses the computational approach taken to develop a sandhi splitter/joiner module and the complexities involved therein. In order to develop sandhi module discussed in this paper Kaccāyana grammar² has been referred to; as it's rules are comprehensive and supported with a lot of examples.

2 Nature of Pāli Sandhi

Words in Pāli language, end in vowel or anusvāra (niggahita). This feature distinguishes it from Samskrita sandhi structure. Pāli sandhis can be divided into internal and external sandhis. Internal sandhis occur within a word and external sandhis are between words. Mainly sandhis are divided based on what pūrvapada (preceding word) ends with and what uttarapada (following word) begins with. They are divided as follows.

2.1 Svarasandhi

When vowels come in proximity as the end of the pūrvapada and the beginning of uttarapada following changes may occur. Say x is the ending vowel of Pūrvapada, y is the beginning vowel of uttarapada.

1. x may get elided, y remains same.

e.g. समेतु + आयस्मा → समेतायस्मा³

¹This research is conducted with the financial assistance from Pāli Institute, Kalburgi. The authors acknowledge the support of Shri Rahul Kharge, Chairman, Pāli Institute, Kalburgi and Prof. Mallepuram G Venkatesh, Hon. Director, Pāli Institute, Kalburgi.

²Tiwari, Laxminarayan and Sharma Birbal (1962) 'Kaccayana vyakarana[Pāli Grammar]', Tara Publications, Varanasi

³सरा सरे लोपं १.२.१

2. x remains same y may get elided.
e.g. चत्तारो + इमे -> चत्तारोमे⁴
3. x gets elided and y could be replaced by asavarṇa vowel
e.g. न + उपेति -> नोपेति⁵
4. x gets elided and y could be replaced by savarṇa long vowel.
सद्धा + इध -> सद्धीध⁶
5. x might get converted to semivowel.
e.g. ते + अस्स -> त्यस्स⁷
6. a consonant may get added between x and y .
e.g. लहु + एस्सति -> लहुमेस्सति⁸
न + इमस्स -> नयिमस्स

2.2 Pakatibhāva

When a word ends with a vowel and is followed by a consonant at the beginning of the uttarapada then both the words remain the same.⁹

e.g. तिण्णो + पारगतो -> तिण्णो पारगतो

When a word ends with a vowel and is followed by a vowel at the beginning of the uttarapada then both the words remain the same.¹⁰

e.g. को + इमं -> को इमं

If the preceding vowel is long it may become short.¹¹

e.g. भोवादी + नाम -> भोवादि नाम

If the preceding vowel is short, it may become long.¹²

e.g. मुनि + चरे =मुनी चरे

2.3 Vyañjanasandhi

In sandhi, if a word ends with a vowel and is followed by a consonant, it is considered as vyañjanasandhi.

e.g. इध + पमादो -> इधप्पमादो¹³

The rule for the above example says, if a word ends with a vowel and is followed by a consonant at the beginning of the following word, the latter gets doubled optionally. This word ‘optionally’ is frequently found in the sūtra or its vṛtti. Following is an example where the doubling of consonant does not happen.

e.g. इध + मोदति -> इध मोदति

⁴वा परो असरूपा १.२.२

⁵क्वचासवण्णं लुत्ते १.२.३

⁶दीघं १.२.४

⁷यमेदन्तस्सादेसो १.२.६

⁸यवमदनतरळा चागमा १.४.६

⁹सरा पकति व्यञ्जने १.३.१

¹⁰सरे क्वचि १.३.२

¹¹स्सं १.३.४

¹²दीघं १.३.३

¹³परद्वेभावो ठाने १.३.६

2.4 Niggahitasandhi

If a word ends with niggahita, followed by a word beginning with a vowel or a consonant, it is considered as niggahita sandhi where niggahita undergoes changes.

e.g. धम्मञ्चरे = धम्मं + चरे.

Here, niggahita changes to ञ according to the rule वगन्त वा वगे १.४.२.

According to this rule, if a vargīya consonant is preceded by niggahita, niggahita gets replaced with अनुनासिक of the same varga. This rule is similar to Pāṇinian rule यरोऽनुनासिकेऽनुनासिको वा ८.४.४५.

3 Computational rules for Sandhi Joiner/Splitter

Pāli sandhi rules stated in the Pāli grammar books of Kacchayana and others are discussed above. Keeping these rules in view, the computational rules for developing module were drawn based on Paninian rules of Sandhi. Following instances were considered:

3.1 Svāra + Savarṇasvāra

A svāra (x) followed by a savarṇa svāra (y), there are five possibilities:

1. lopa of x and y remains.
2. lopa of x and y gets elongated.
3. anusvāra or y/v/m/d/n/t/r/L may be inserted between x and y.
4. prakṛtibhāva.

For e.g अ + अ -> अ/आ/अनुस्वार insertion/य,व,द,म,त,र,ळ insertion/remain the same

तत्र + अयं ->

- >तत्रयं (replaced with अ) (1)
- >तत्रायं (replaced with आ) (2)
- >तत्रं अयं (अनुस्वार insertion) (3)
- >तत्ररयं (र insertion) (4)
- >तत्र अयं (remain the same) (5)

Outputs (3), (4) and (5) are not seen in sample gold data. Hence, these options can be hidden in the display.

3.2 Svāra + Asavarṇasvāra

A svāra (x) followed by an asavarṇa svāra (y), there are seven possibilities:

1. x remains and lopa of y.
2. lopa of x and y gets guṇa.
3. all rules of section 3.1

For e.g. आ + इ -> इ/आ/ई/अनुस्वार insertion/remain the same

लता + इव ->

- >लताव (1)
- >लतेव (2)
- >लतिव (3)
- >लतीव (4)
- >लतां इव (5)
- >लतामिव (6)

->लता इव (7)

Outputs (5), (6) and (7) are not seen in sample gold data. Hence, these options can be hidden in the display.

3.3 svāra + vyañjana

A svāra (x) followed by vyañjana (y), there are six possibilities:

1. elongation of x and y remains.
2. x may get replaced with अ and ओ and y remains.
3. anusvāra may be inserted between x and y.
4. prakṛtibhāva.

इ + च ->ई/अनुस्वार insertion/इ ->अ/इ ->ओ/remain the same e.g.1 मुनि + चरे ->

- >मुनी चरे (1)
- >मुनो चरे (2)
- >मुन चरे (2)
- >मुनिच्चरे (3)
- >मुनिं चरे (4)
- >मुनि चरे (5)

e.g.2 इध + पमादो ->

- >इधा पमादो (1)
- >इधो पमादो (2)
- >इध पमादो (2)
- >इधप्पमादो (3)
- >इधं पमादो (4)
- >इध पमादो (5)

Outputs (4) and (5) are not seen in sample gold data. Hence, these options can be hidden in the display.

3.4 niggahita(anusvāra) + svāra

A niggahita (x) followed by svāra (y), there are four possibilities:

1. lopa of x, elongation of upadha and y remains.
2. lopa of x and y remains.
3. lopa of x and म् / ढ् may be inserted between x and y.
4. x remains and lopa of y
5. prakṛtibhāva.

For e.g. anusvāra + अ ->elision of anusvāra and elongation of upadha/elision of anusvāra/elision of अ /insertion of म्

तासं + अहं ->

- >तासाहं (1)
- >तासहं (2)
- >तासमहं (3)
- >तासं हं (4)
- >तासं अहं (5)

Outputs (4) and (5) are not seen in sample gold data. Hence, these options can be hidden in the display.

3.5 niggahita(anusvāra) + vyañjana

A niggahita (x) followed by vyañjana (y), there are three possibilities:

1. x gets replaced with nasal of the same varga.
2. lopa of x and y remains.
3. lopa of x and ण् /द् may be inserted between x and y.
4. prakṛtibhāva.

For e.g. anusvāra + च -> elision of anusvāra/anusvāra ->nasal of same varga/remain the same धम्मं + चरे ->

- >धम्मञ्चरे (anusvāra to nasal of same varga) (1)
- >धम्मचरे (elision of anusvāra) (2)
- >धम्मं चरे (remain the same) (3)

Outputs (2) and (3) are not seen in sample gold data. Hence, these options can be hidden in the display.

3.6 Apavāda rules

pūrvapada (x) and is followed by vyañjana (y):

3.6.1 Apavāda 1

1. if x = पुथ, last letter is replaced by उ and y remains.
2. if x = पुथ, last letter is replaced by उ and y may get doubled.
3. prakṛtibhāva. पुथ + भूतं -> पुथुभूतं
पुथ + जनो -> पुथुज्जनो

3.6.2 Apavāda 2

1. if x = अव, x is replaced with औ and y remains.
2. prakṛtibhāva. अव + नद्धा -> औनद्धा
अव + नद्धा -> अवनद्धा

3.6.3 Apavāda 3

1. if x = पति, x is replaced with पटि and y remains
पति + हञ्जति -> पटिहञ्जति

pūrvapada (x) and is followed by vyañjana (y):

3.6.4 Apavāda 4

1. if x = पा, last letter of x is shortened and ण् is inserted between x and y.
2. prakṛtibhāva. पा + एव -> पणोव
पा + एव -> पाणव

3.6.5 Apavāda 5

1. if x = अभि, x is replaced with अब्भ् and y remains
अभि + उदीरितं -> अब्भुदीरितं

3.6.6 Apavāda 6

1. if x = अधि, x is replaced with अज्झ् and y remains
अधि + ओकासो -> अज्झोकासो

3.6.7 Apavāda 7

1. if $x = \text{अभि/अधि}$ and $y = \text{इ}$, lopa of last letter of x .
2. if $x = \text{अभि}$ and $y = \text{इ}$, lopa of last letter of x .
3. Apavāda 5 and 6 are applicable here.
अभि + इज्झितं -> अभिज्झितं
अभि + इज्झितं -> अब्भिज्झितं
अधि + ईरितं -> अधीरितं
अधि + ईरितं -> अज्झीरितं

3.6.8 Apavāda 8

1. if $x = \text{अति}$, and $y = \text{इ}$, lopa of last letter of x .
अति + ईरितं -> अतीरितं

3.6.9 Apavāda 9

1. if $x = \text{पति}$, x is replaced with पटि , lopa of last letter of x and y remains.
पति + अग्नि -> पटग्नि

From sutra सरा पकति व्यञ्जने १.३.१ and सरे क्वचि १.३.२ together, it can be derived that, if ending vowel of a word comes in proximity of beginning vowel/consonant of the following word, both remain the same (prakṛtibhāva). Therefore the output of this instance need not be shown in the display. Because every instance of sandhi, prakṛtibhāva can happen. Similarly, sūtra निग्गहितञ्च १.४.८ indicates insertion of anusvāra for every sandhi instance. Hence output here also can be selectively shown.

4 Complexities Involved

While drawing rules for Pāli sandhi computation, the following complexities were encountered. In the first place, we notice words like क्वचा, वा which means sometimes or optional in many sutras. That makes most of the sandhis optional or having multiple results based on the situation of x and y . Below are some examples to demonstrate the complexities.

4.1 Occurrence of words क्वचा and वा

Majority of sutras i.e. out of 41 kaccāyana sandhi sutras almost 27 sutras have क्वचा and वा in sutra itself or the vṛtti. For e.g वमोदुदन्तानं rule क्वचा is in the vṛtti.¹⁴ This gives rise to multiple outputs for a given instance when generated computationally. In the case of Sanskrit this ambiguity is mostly fixed by rules themselves. If there are exceptions, they are grouped and gaṇa information is provided. In Pāli, one has to depend heavily on literature to get the forms that are used rather than those which can be generated computationally.

4.2 Inconsistency in examples from literature

Following are examples from piṭakasahitya:

भिकखवे+इति => भिकखवेति
ए+ इ => ए + _ वा परो असरूपा

आवुसो +इति => आवुसोति
ओ+ इ => ओ + _ वा परो असरूपा

It is observed that वा परो असरूपा rule is followed in the above examples. This rule is optional but it is applied most of the time wherever dissimilar vowels come in proximity of each other in

¹⁴Pg 19 Tiwari, Laxminarayan and Sharma Birbal(1962) 'Kaccayana vyakarana[Pāli Grammar]', Tara Publications, Varanasi

sandhi.

But in the following example from the same text, though dissimilar vowels are in proximity of each other, it is seen that सरा सरे लोपं and दीघं are applied. So this seems to be an exception to the above rule.

च + इध => चीध
अ + इ => _ + इ सरा सरे लोपं
_ + इ => _ + ई दीघं

Whereas in the examples below सरा सरे लोपं and दीघं are followed where similar vowels are in close contact in sandhi.

भुञ्जामि+इति => भुञ्जामीति इ+ इ => _ + इ सरा सरे लोपं
_ + इ => _ + ई दीघं Another example from the same text
पमुच्छति+इति => पमुच्छतीति
न+अत्थि => नत्थि¹⁵ सरा सरे लोपं

Here elongation of vowel has not occurred. Therefore even from literature, joining or splitting has to be done with caution.

4.3 Ambiguous Rules for Insertion of Letters

Insertion of letters in Pāli sandhi is ambiguous. For e.g. application of the rule called यवमदनतरळा चागमा. This rule says if uttarapada begins with svāra then optionally य/व/म/द/न/त/र/ळ can be inserted. In the examples given in the vṛtti:

सम्मा + अज्जा -> सम्मदज्जा-> आ + अ -> द insertion
भन्ता + उदिक्खति -> भन्तावुदिक्खति -> आ + उ -> व insertion
अज्ज + अग्गे -> अज्जतग्गे-> अ + अ -> त insertion
अत्त + अत्थभिज्जाय -> अत्तदत्थभिज्जाय-> अ + अ -> द insertion

It is observed that for a given instance, the letter inserted is different for a similar condition. Extracting rules from such sūtrās is difficult.

4.4 Multiple Possibilities while Splitting

Multiple sūtrās are available for splitting the same instance.

For e.g. लतेव can be split as
लता + एव — 1 सरा सरे लोप
लता + इव — 2 क्वचासवण लुत्ते

Above example shows that the split has to be context-based.

In Sanskrit लतेव can be split in only one way i.e. लता + इव and this can be context-independent.

5 Sandhi Joiner

Sandhi Joiner was developed applying the rules enumerated in the previous section. The input to the tool is pūrvapada and uttarapada. The result is all possible combined words based on the rules that are applicable to a given instance. It also indicates the respective rules which are applied to get that particular output. Sandhi Joiner has three modules - svarasandhi, vyañjanasandhi, and niggahitasandhi. Pseudocode for the tool is given section 4.1. Flow chart is given below. The kaccāyana rules used in the respective modules are listed in A Appendix - 1. The screenshots are attached in B Appendix-2. The computational module for the Sandhi Splitter is the reverse of sandhi joiner. The work for this module is under process.

5.1 Pseudocode

Begin

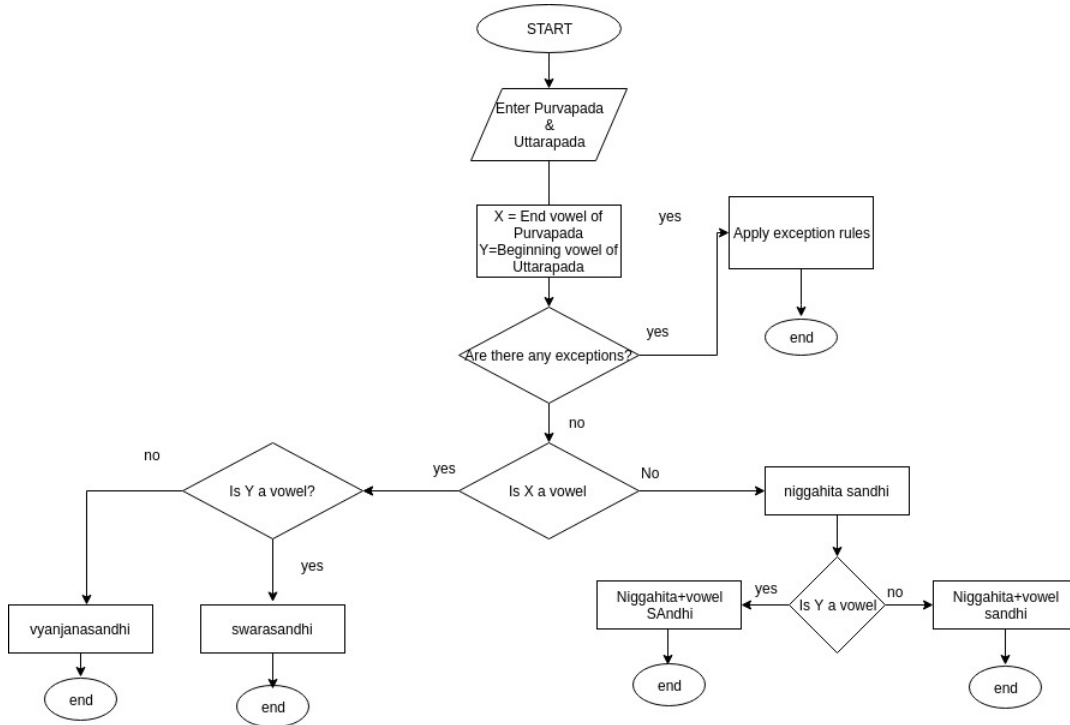
Input pūrvapada and uttarapada

If exceptions exist then

 Derive required output

¹⁵Muller E.(1962) 'A simplified grammar of Pāli Language', Trubner and company, London

Display output
 Exit program
 Assign X as ending varna(character) of pūrvapada
 Assign Y as beginning letter of uttarapada
 If X and Y are vowels then
 Go to svarasandhi module
 Derive required output
 Display output
 exit the program
 If Y is vyañjana then
 Go to vyañjanasandhi module
 Derive required output
 Display possible output
 If X is niggahita then
 Go to niggahitasandhi module
 If Y is svara then
 Go to niggahita-svara module
 Derive required output
 Display output
 exit the program
 If Y is vyañjana then
 Go to niggahita-vyañjana module
 Derive required output
 Display output
 exit the program
 End



5.2 Statistics

For validating the tool, 398 sandhi examples are collected from various Pāli grammar and other texts. This data was put through the sandhi tool and compared with gold data. Following are the statistics of the output.

Total number of words	398
Total number of outputs matching atleast one gold data	356
svarasandhi	158
vyañjanasandhi	82
niggahitasandhi	158
apavāda	13
single output matching gold data	26
two outputs matching gold data	13
three outputs matching gold data	130
five outputs matching gold data	19
six outputs matching gold data	14
seven outputs matching gold data	76
eight outputs matching gold data	58
Nine outputs matching gold data	19
outputs not matching gold data	44

By examining the statistics, we notice that svarasandhi and niggahitasandhi are equal in number. Out of 398 words, 84% outputs had at least one output matching with gold data. We observed that multiple outputs are more in case of svarasandhi. Since our focus is on the complexities of sandhi rules, limited examples are taken for validation. More sandhi data will be analyzed later.

6 Scope for Future work

1. More examples from Pali literature have to be collected to validate the tool.
2. Exhaustive Statistical study of the Pali literature has to be undertaken to decide which sandhi rule is frequently applied to a given instance.
3. Pruning the outputs based on statistics.
4. Integrating with a dictionary to reduce multiple outputs.

7 Conclusion

Making a full-fledged sandhi splitter/joiner is a complex process due to the ambiguous sandhi rules. As seen by the results of Sandhi Joiner, for a given instance, there is a probability of multiple outputs. This is because of the nature of Pāli words and the complex nature of the grammatical rules. With the understanding of the nature of language, to prune the outputs, a wider study of literature is required.

References

- Tiwari, Laxminarayan and Sharma Birbal (1962) ‘Kacchayana vyakarana[Pāli Grammar]’, Tara Publications, Varanasi
- Duroiselle Charles (1921) A practical grammar of the Pāli Language. British Burma Press, Rangoon
- Muller E. (1962) ‘A simplified grammar of Pāli Language’, Trubner and company, London
- Childers, Robert Caesar (1875) “A dictionary of Pāli language”, Trubner and co., London
- Davids, Rhys and William Stede (1921-25) “Pāli – English Dictionary”, Pāli Text Society.
- D’Alwis, James (1863) “Introduction to kachchayanans grammar of the Pāli language”, Colombo

Kashyap, Jagadish Bhikshu (1940) “Pāli Mahavyakarana”, Mahabodhi Sabha, Saranath, Banaras.

Saddhatissa, (1949) “Saral Pāli Shiksha”, Mahabodhi society, Saranath, Banaras

Silananda, Venerable U “Pāli roots in Saddaniti”, e-book.

Tungar, Na Va (1939) “Pāli Bhasha Pravesha”, Samarth Bharat, Pune.

A Appendix - 1

Kaccāyana rules used in various sandhi modulewise

A.1 Case 1 : vowel + vowel

सरा सरे लोपं	१.२.१
वा परो असरूपो	१.२.२
क्वचासवण्ण लुत्ते	१.२.३
दीघं	१.२.४
पुब्बो च	१.२.५
यमेदन्तस्सादेसो	१.२.६
वमोदुदन्तानं	१.२.७
सब्बो चन्ति	१.२.८
दो धस्स च	१.२.९
इवण्णो यन्न वा	१.२.१०
एवादिस्स रि पुब्बो च रस्सो	१.२.११

A.2 Case 2 : vowel + consonant

सरा पकति व्यञ्जने	१.३.१
दीघं	१.३.३
रस्सं	१.३.४
लोपश्च तत्राकारो	१.३.५
परद्वेभावो ठाने	१.३.६
वग्गे घोसाघोसानं ततियपठमा	१.३.७

A.3 Case 3 : niggahita + vowel

मदा सरे	१.४.५
यवमदनतरळा	१.४.६
क्वचि लोप	१.४.९
व्यञ्जने च	१.४.१०
परो वा सरो	१.४.११
व्यञ्जनो च विसञ्जोगो	१.४.१२

A.4 Case 4 : niggahita + consonant

अं व्यञ्जने निग्गहीतं	१.४.१
वगन्त वा वग्गे	१.४.२

एहेय्यं	१.४.३
सये च	१.४.४

A.5 Case 5 : special sandhis

क्वचि ओ व्यञ्जने	१.४.७
निग्गहितञ्च	१.४.८
गो सरे पुथुस्सागमो क्वचि	१.५.१
पास्स चन्तो रस्सो	१.५.२
अब्भो अभि	१.५.३
अज्झो अधि	१.५.४
ते न वा इवण्णे	१.५.५
अतिस्स चन्तस्स	१.५.६
क्वचि पटि पतिस्स	१.५.७
पुथुस्स व्यञ्जने	१.५.८
ओ अवस्स	१.५.९
अनुपदिट्ठान वुत्तयोगतो	१.५.१०

B Appendix - 2

Screenshots of sample input and sample outputs are given below.

Figure 1: Input

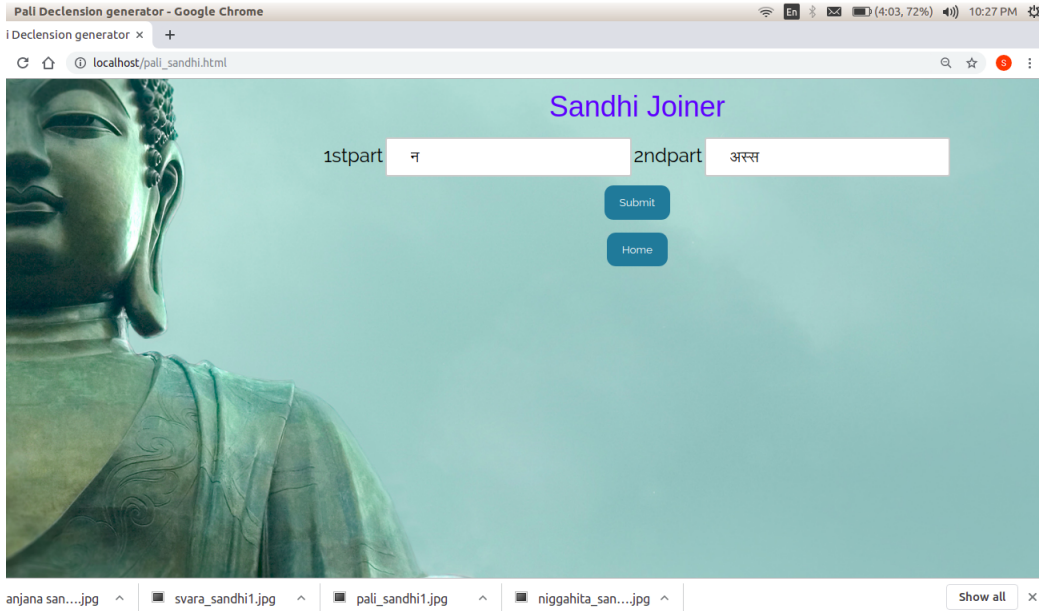


Figure 2: Output

swati-inspron-5559/cgi-bin/pali_sandhi.cgi - Google Chrome

swati-inspron-5559/cgi-b... x +

Not secure | swati-inspron-5559/cgi-bin/pali_sandhi.cgi

in first point na in second point na

न अस्स

निग्गहीतञ्च	नअस्स
सरा सरे लोप	नस्स
दीघं	नास्स
यवमदनतरळा	नतस्स
पुथस्सु व्यञ्जने	नुस्स
सरे क्वचि	न अस्स

Home

vyanjana san...jpg ^ svara_sandhi1.jpg ^ pali_sandhi1.jpg ^ niggahita_san...jpg ^ Show all x