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 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 grammar[1] 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 pūvapada 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 –> समेतायस्मा[2]

2. x remains same y may get elided.
   e.g. चत्तारो + इमे –> चत्तारोमे[3]

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

[2] सरा सरे लोपं १.२.१

[3] वा परो असरूपा १.२.२

3. x gets elided and y could be replaced by asavarṇa vowel
   e.g. न + उपेति –> नोपेति[4]

4. x gets elided and y could be replaced by savarṇa long vowel.
   सद्धा + इध –> सद्धीध[5]

5. x might get converted to semivowel.
   e.g. ते + अस्स –> त्यस्स[6]

6. a consonant may get added between x and y.
   e.g. लहु + एस्सति –> लहुमेस्सति[7]
   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.[8]
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.[9]
e.g. को + इमं –> को इमं

If the preceding vowel is long it may become short.[10]
e.g. भोवादी + नाम –> भोवादि नाम

If the preceding vowel is short, it may become long.[11]
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 –> इधप्पमादो[12]

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

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.

---

[4]क्वचासवण्णं लुत्ते १.२.३
[5]दीघं १.२.४
[6]यमेदन्तस्सादेसो १.२.६
[7]यवमदनतरळा चागमा १.४.६
[8]सरा पकति व्यञ्जने १.३.१
[9]सरे क्वचि १.३.२
[10]रस्सं १.३.४
[11]दीघं १.३.३
[12]परद्वेभावो ठाने १.३.६

e.g. धम्मञ्चरे = धम्मं +चरे.

Here, niggahita changes to ञ् according to the rule वग्गन्त वा वग्गे 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 Svara + Savarṇasvara

A svara (x) followed by a savarṇa svara (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 Svara + Asavarṇasvara

A svara (x) followed by an asavarṇa svara (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 svara + vyañjana

A svara (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) + svara

A niggahita (x) followed by svara (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 /द् 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 म
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 म् /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 ओ and y remains.
2. prakṛtibhāva. अव + नद्धा –> ओनद्धा
   अव + नद्धा –> अवनद्धा

3.6.3 Apavāda 3

1. if x = पति, x is replaced with पटि and y remains
   पति + हञ्ञति –> पटिहञ्ञति

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

3.6.4 Apavāda 4

1. if x = पा, last letter of x is shortened and ग् is inserted between x and y.
2. prakṛtibhāva. पा + एव –> पगेव
   पा + एव –> पाएव

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 = अभि/अधि and y = इ, lopa of last letter of x.

2. if x = अभि and y = 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 = अति, and y = इ, lopa of last letter of x.
   अति + ईरितं –> अतीरितं

3.6.9 Apavāda 9

1. if x = पति, x is replaced with पटि, lopa of last letter of x and y remains.
   पति + अग्गि –> पटग्गि

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, prakrutibhāva 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. [13] 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:

भिक्खवे+इति => भिक्खवेति
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

[13] Pg 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 => नत्थि[14] 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 svara then optionally य/व/म/द/न/त/र/ळ can be inserted. In the examples given in the vṛtti:

सम्मा + अञ्ञा –> सम्मदञ्ञा–> आ + अ –> द insertion
भन्ता + उदिक्खति –> भन्तावुदिक्खति –> आ + उ –> व insertion
अज्ज + अग्गे –> अज्जतग्गे–> अ + अ –> त insertion
अत्त + अत्थभिञ्ञाय –> अत्तदत्थभिञ्ञाय–> 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 sutrās is difficult.

4.4 Multiple Possibilities while Splitting

Multiple sutrās are available for splitting the same instance.

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

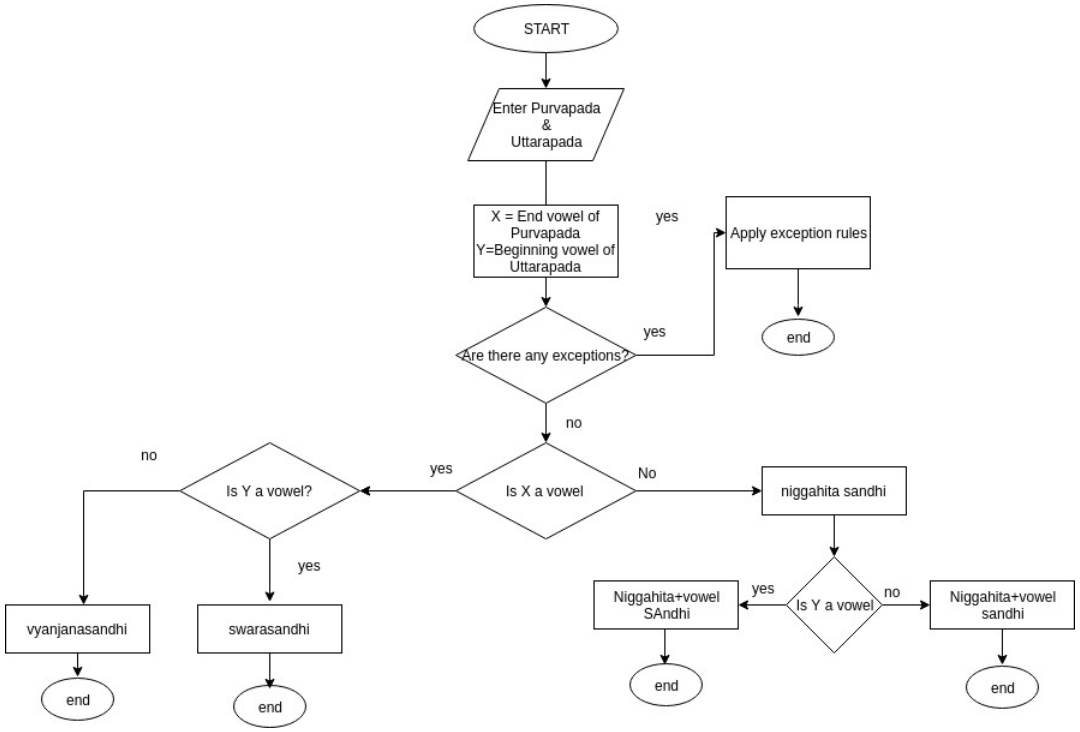
Sandhi Joiner was developed applying the rules enumerated in the previous section. The input to the tool is pūrvapada and uttarapada. The result is all possible combined words based on the rules that are applicable to a given instance. It also indicates the respective rules which are applied to get that 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

[14]Muller 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 kachchayanas 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 असरुपो	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 विसञ्ञोगो	१.४.१२

A.4 Case 4 : niggahita + consonant

अं व्यञ्जने निग्गहीतं	१.४.१
वग्गन्त 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

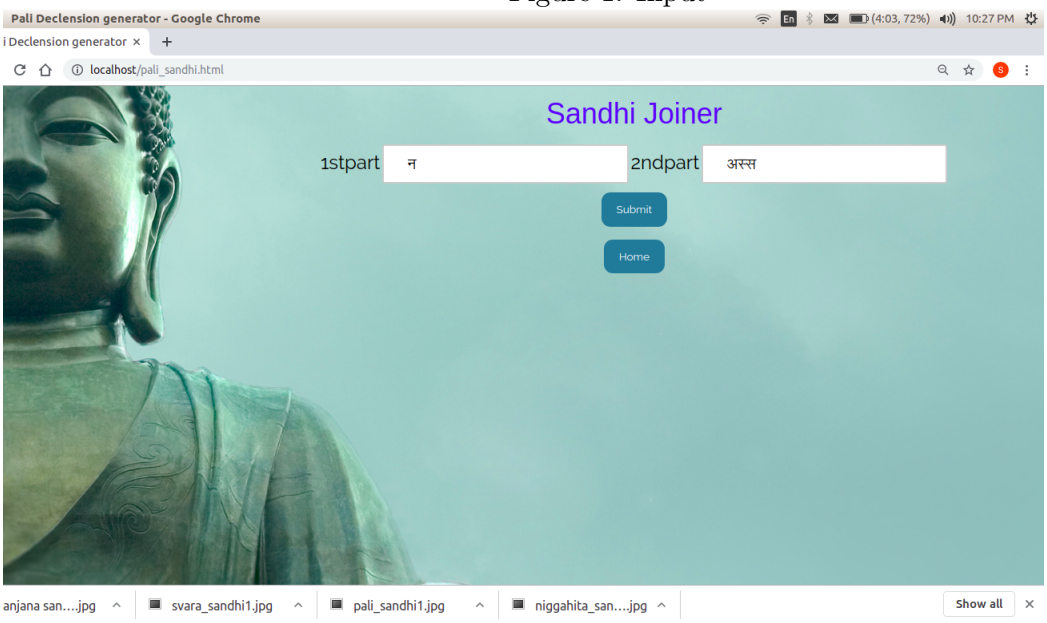


Figure 2: Output