A New Computational Schema for Euphonic Conjunctions in Sanskrit Processing

Rama N. and Meenakshi LAKSHMANAN

1 Department of Computer Science, Presidency College
Chennai 600 005, India

2 Department of Computer Science, Meenakshi College for Women
Chennai 600 024, India
and
Research Scholar, Mother Teresa Women’s University
Kodaikanal 624 101, India

Abstract
Automated language processing is central to the drive to enable facilitated referencing of increasingly available Sanskrit E-texts. The first step towards processing Sanskrit text involves the handling of Sanskrit compound words that are an integral part of Sanskrit texts. This firstly necessitates the processing of euphonic conjunctions or sandhi-s, which are points in words or between words, at which adjacent letters coalesce and transform.

The ancient Sanskrit grammarian Pāṇini’s codification of the Sanskrit grammar is the accepted authority in the subject. His famed sūtra-s or aphorisms, numbering approximately four thousand, tersely, precisely and comprehen sively codify the rules of the grammar, including all the rules pertaining to sandhi-s.

This work presents a fresh new approach to processing sandhi-s in terms of a computational schema. This new computational model is based on Pāṇini’s complex codification of the rules of grammar. The model has simple beginnings and is yet powerful, comprehensive and computationally lean.

Keywords: Sanskrit, euphonic conjunction, sandhi, linguistics, Panini, aphorism, sutra.

1. Introduction
The recognition of Sanskrit as a highly phonetic language as also one with an extensively codified grammar [1], is widespread. The very name Sanskrit (Sanskrit) means "language brought to formal perfection". That the Backus-Naur Form used in the specification of formal languages, has now come to be popularly known as the Pāṇini-Backus Form [8, 9], bears ample testimony to this fact.

Sanskrit E-texts are being increasingly made available for reference in repositories such as the Göttingen Register of Electronic Texts in Indian Languages (GRETI L) [11]. Now the essential first step towards language processing of such Sanskrit E-texts is to develop efficient algorithms and tools to handle segmentation in Sanskrit compound words that are an integral part of Sanskrit texts. This firstly necessitates the processing of sandhi-s or euphonic conjunctions.

1.1 Unicode Representation
The Unicode (UTF-8) standard is what has been adopted universally for the purpose of encoding Indian language texts into digital format. The Unicode Consortium has assigned the Unicode hexadecimal range 0900 - 097F for Sanskrit characters.

All characters including the diacritical characters used to represent Sanskrit letters in E-texts are found dispersed across the Basic Latin (0000-007F), Latin-1 Supplement (0080-00FF), Latin Extended-A (0100-017F) and Latin Extended Additional (1E00 – 1EFF) Unicode ranges.

The Latin character set has been employed in this work to represent Sanskrit letters as E-text. Moreover in this paper, any Sanskrit text except the names of people is given in italics. As such, variables such as x, y and z are not italicized as per the norm.

1.2 The Basis of the Work
Pāṇini, the sage and scholar dated by historians in the fourth century BC or earlier, codified the rules of the Sanskrit language based on both the extant vast literature as well as the language in prevalent use at the time. His magnum opus, the Aṣṭādhyāyī, which literally means ‘work in eight chapters’, is regarded by all scholars as the ultimate authority on Sanskrit grammar. In four parts each, these eight chapters comprise nearly four thousand sūtra-s or aphorisms, terse statements in Sanskrit. This grammar-codification of Pāṇini is perhaps unparalleled, for it is terse and yet comprehensive, complex yet precise. Intensive study, taking recourse to authoritative commentaries authored by adroit grammarians, is required to get a grasp of the work.

Many commentaries on the Aṣṭādhyāyī, such as Sage Patañjali’s Mahābhāṣya are available and held as authentic
and comprehensive. One such authoritative commentary with a neat, topic-wise classification of Pāṇini’s aphorisms, is the Siddhānta-kaumudi [2] written in the seventeenth century by the Sanskrit grammarian, Bhaṭṭoji Dīkṣīta. The most important of these aphorisms were later extracted and compiled into the Laghu-siddhānta-kaumudi [10] by the scholar Varadarāja.

It is accepted among Sanskrit scholars that any exploratory work on Sanskrit grammar must necessarily have the aphorisms of Pāṇini as its basis, optionally taking recourse to any of the authoritative commentaries. This work on euphonic conjunctions is also based directly on Pāṇini’s aphorisms, and not on secondary or tertiary sources of information. The Siddhānta-kaumudi of Bhaṭṭoji Dīkṣīta, famed and accepted amongst scholars as an unabridged, comprehensive compendium of the entire Astādhyāyī, has been studied in the original Sanskrit, and the euphonic conjunctions dealt with in it form the basis of this work. The Laghu-siddhānta-kaumudi was also initially consulted for insights.

1.3 The Māheśvara aphorisms - the backbone of Pāṇini’s code

The Māheśvara aphorisms, said to have come from the beats of a special drum called ‘damaru’ (hourglass drum) held in the hand of Lord Mahesvara (a form of God in the Hindu pantheon), are a set of aphorisms containing the letters of the Sanskrit alphabet in a certain sequence. These aphorisms form the basis of Pāṇini’s composition of his grammar aphorisms. The Māheśvara aphorisms are fourteen in number and are listed below:

1. a-i-u-n 2. r-l-k 3. e-o-ṅ 4. ai-a-u-c 5. ha-ya-va-ra-t 6. la-ṅ 7. ṇa-ma-ṇa-na-m 8. jha-bha-ṅ 9. gha-tha-da-s 10. ja-ba-ga-da-da-s 11. kha-pha-cha-tha-ca-ṭa-ta-v 12. ka-pa-ṛ 13. sa-sa-a-r 14. ha-l

The last letter in each of the above aphorisms is only a place-holder and is not counted as an actual letter of the aphorism. The first four aphorisms list the short forms of all the vowels, while the rest list the consonants. It must be noted that the letter ‘a’ added to each of the consonants is only to facilitate pronunciation and is not part of the consonant proper.

2. The Problem

Sandhi-s in Sanskrit are points in words or between words, at which adjacent letters coalesce and transform. This is a common feature of Indian languages and is particularly elaborately dealt with and used in Sanskrit. The transformations that apply are commonly categorized into four:

1. āgama – addition of an extra letter or set of letters
2. ādeśa – substitution of one or more of the letters
3. lopa – dropping of a letter
4. prakṛtiḥbhāva – no change

(The last is considered a transformation in the language and has therefore been listed above. However, it may be ignored for practical purposes and is hence not covered in this work.)

There are close to seventy aphorisms of Pāṇini that deal with sandhi-s. These aphorisms lay out the rules for the above transformations, giving the conditions under which certain letters combine with certain others to give particular results.

The challenge is to develop a computational algorithm to handle the entire range of sandhi-s. Such a computational algorithm would be useful to generate various word forms of a given Sanskrit word through the application of sandhi rules. Though this task is not difficult for a scholar of Sanskrit with a thorough knowledge of the Pāṇinian system, it is certainly a computationally non-trivial task, given the complexity and number of rules.

Existing methods of sandhi processing, be they methods to form compound words or even to try to split them, seem to be based on a derived understanding of the functioning of euphonic conjunctions, and usually go the finite automata-HMM-artificial intelligence way [3-7, 12]. However, the present work directly codifies Pāṇini’s rules as is, recognizing that Pāṇini’s codification of the grammar is based on the Māheśvara aphorisms that in turn lay out the letters of the alphabet in a non-trivial order. This work presents one novel method of directly representing Pāṇini’s sandhi rules. It presents, on this basis, a mathematical formulation of a new approach to solving the non-trivial problem of handling euphonic conjunctions.

3. The Approach

To take advantage of the ordering of letters of the alphabet given in 2.3 above, we assign values to each letter in the Sanskrit alphabet, sticking to the order in the Māheśvara aphorisms rather than to the commonly adopted ordering of the letters. Thus, we have the assignment of values for the letters of the alphabet shown in Table 1.

| Letter | Value | Letter | Value | Letter | Value |
|--------|-------|--------|-------|--------|-------|
| a      | 1     | l      | 18    | ph     | 35    |
| ã      | 2     | ñ      | 19    | ch     | 36    |
| i      | 3     | m      | 20    | ṭh     | 37    |
| ì      | 4     | ñ      | 21    | th     | 38    |
| u      | 5     | n      | 22    | ̄c      | 39    |
| û      | 6     | ṇ      | 23    | t      | 40    |
| r      | 7     | jh     | 24    | t      | 41    |
| ñ      | 8     | bh     | 25    | g      | 42    |
| l      | 9     | gh     | 26    | ṇ      | 43    |
| e      | 10    | dh     | 27    | s      | 44    |
| o      | 11    | dh     | 28    | s      | 45    |

Table 1 : Values for the letters of the Sanskrit alphabet

IJSI International Journal of Computer Science Issues, Vol. 5, 2009
Further, the letters are clubbed into various types as given below:

1. vowels: 1 – 13
2. consonants: 14 – 47
3. semi-vowels: 15 – 18
4. mutes: 19 – 47
5. nasals: 19 – 23
6. non-nasal mutes: 24 - 47
7. soft consonants: 24 – 33
8. hard consonants: 34 - 46
9. column 1: 39 – 43
10. column 2: 34 – 38
11. column 3: 29 – 33
12. column 4: 24 – 28
13. sibilants: 44 – 46
14. aspirate: 14 and 47
15. anuvāra: 48
16. visarga: 49
17. avagraha: 50 (replacement for the first vowel)
18. ru: 51 (denotes the letter r but is handled differently)
19. gutturals: 42, 34, 31, 26, 21
20. palatals: 39, 36, 29, 24, 19
21. cerebrals: 40, 37, 32, 27, 22
22. dentals: 41, 38, 33, 28, 23
23. labials: 43, 35, 30, 25, 20

A rule is the name we use for letter-level conjunctions such as the following of the savarnadīrgha type: \(a+ā = ā\) where the symbol ‘+’ denotes adjacency and the term on the right of the ‘=’ symbol is the resultant term that has to be either substituted for or added to ones on the left. (In the case of this particular sandhi, the term on the right is the single substitute for both terms on the left.) Substituting values of letters from Table 1, this would translate into \(1 + 2 = 2\).

Each sandhi may have more than one governing aphorism that specifies its functioning. Each such aphorism for every sandhi type in turn expands into a series of ‘rules’ as defined above. In this work, each and every rule for each aphorism under each of the major twenty three sandhi types were listed. Further, an aphorism would specify if an addition, deletion or substitution would have to be made. In accordance with this, a further cataloguing of aphorisms into five categories was done.

If we denote a sandhi rule as \(x + y = z\) where variables \(x\) and \(y\) denote the values of single letters joining together to yield a resultant \(z\), then we have the following categorizations depending on both the characteristics of \(z\) and on what we actually do with it:

- \(C_1\): replace \(x\) and \(y\) with single-letter or multi-letter \(z\)
- \(C_2\): replace \(x\) with single-letter or multi-letter \(z\)
- \(C_3\): replace \(y\) with single-letter \(z\)
- \(C_4\): add single-letter \(z\)
- \(C_5\): drop \(x\)

Table 2 gives the summary of the numbers involved in this scenario. It must be noted that in practice, some aphorisms have to be combined or handled in two different ways to yield sets of rules, and hence what may seem to be a discrepancy in the number of aphorisms shown in the table and the number of rows shown for the rules of that aphorism, is no real discrepancy at all.

As can be seen, there are close to 2500 individual rules involved, even with considering only the major sandhi-s. Tabulation of these rules in terms of \(x\), \(y\) and \(z\) for the categories and then tabulation of the corresponding values as per Table 1 were done.

Table 2: Summary of the number of Sanskrit sandhi aphorisms and rules

| # | Sandhi Type | No. of sātivas | Categories | No. of Rules |
|---|-------------|--------------|------------|--------------|
| 1 | yanādēsa     | 1            | \(C_1\)    | 74           |
|   |             |              | \(C_2\)    | 74           |
| 2 | ayāyāvāvādēsa | 4            | \(C_3\)    | 50           |
|   |             |              | \(C_4\)    | 50           |
| 3 | guṇa        | 2            | \(C_5\)    | 2            |
|   |             |              |            | 3            |
| 4 | vṛddhi      | 3            |            | 18           |
|   |             |              |            | 18           |
| 5 | pararāṣa    | 1            |            | 10           |
|   |             |              |            | 10           |
| 6 | savarnadīrgha | 1          |            | 15           |
| 7 | pārvapāṇa    | 1            |            | 2            |
|   |             |              |            | 2            |
| 8 | avanādēsa   | 1            |            | 13           |
|   |             |              |            | 13           |
| 9 | tūgāgama    | 4            |            | 5           |
|   |             |              |            | 2            |
| 10 | jaśva       | 2            |            | 23           |
|    |             |              |            | 23           |
| 11 | satva       | 2            |            | 230          |
|    |             |              |            | 230          |
|    |             |              |            | 138          |
|    |             |              |            | 138          |
| 12 | anuvāra     | 5            |            | 34           |
|    |             |              |            | 34           |
|    |             |              |            | 24           |
|    |             |              |            | 24           |
|    |             |              |            | 1           |
|    |             |              |            | 1           |
|    |             |              |            | 3           |
|    |             |              |            | 3           |
| 13 | dhudāgama   | 2            |            | 2           |
|    |             |              |            | 2           |
| 14 | namudāgama  | 1            |            | 195          |
|    |             |              |            | 195          |
| 15 | ścūtra      | 2            |            | 36           |
|    |             |              |            | 36           |
|    |             |              |            | 31           |
|    |             |              |            | 31           |
| 16 | śṭutra      | 3            |            | 6           |
|    |             |              |            | 6           |
| 17 | anuvāsikā   | 1            |            | 160          |
| 18 | cartva      | 1            |            | 312          |
|    |             |              |            | 312          |
| 19 | parasaśvarna | 3            |            | 29           |
|    |             |              |            | 29           |
|    |             |              |            | 5           |
|    |             |              |            | 5           |
| 20 | pārvavasvarna| 1           |            | 20           |
|    |             |              |            | 20           |
| 21 | chaśva      | 1            |            | 340          |
|    |             |              |            | 340          |
| 22 | visarga     | 2            |            | 13           |
|    |             |              |            | 13           |
| 23 | svādī       | 5            |            | 66           |
|    |             |              |            | 66           |
|    |             |              |            | 13           |
|    |             |              |            | 13           |
|    |             |              |            | 132          |
|    |             |              |            | 132          |
|    |             |              |            | 33           |
|    |             |              |            | 33           |
|    |             |              |            | 33           |
|    |             |              |            | 33           |
| TOTAL |    | 49           |            | 89           |
|    |             | 1320         |            | 1439         |
|    |             | 1320         |            | 397          |
|    |             |              |            | 211          |
|    |             |              |            | 277          |
|    |             |              |            | 2413         |

Careful observations based on a thorough understanding of the domain and classification of the input conditions, yielded the equations presented in this work.
We define the general binary operators \( \oplus_1, \oplus_2, \oplus_3, \oplus_4 \) and \( \oplus_5 \) for the categories \( C_1, C_2, C_3, C_4 \) and \( C_5 \) respectively, as follows:

\[
\begin{align*}
C_1: \oplus_1(x, y) &= z = z_1 \\
C_2: \oplus_2(x, y) &= z = z_2y \\
C_3: \oplus_3(x, y) &= z = xz_3 \\
C_4: \oplus_4(x, y) &= z = xz_4y \\
C_5: \oplus_5(x, y) &= z = y
\end{align*}
\]

where each of \( z_1, z_2, z_3, z_4 \) is to be calculated. Now we introduce the second and third subscripts for the above general operators as follows: the general operator \( \oplus_{i,j}(x, y) \) is derived from \( \oplus_i \) and signifies the operator applying to aphorism number \( j \) of Category \( C_i \); the specialized operator \( \oplus_{i,j,k}(x, y) \) is derived from the operator \( \oplus_{i,j} \) and appertains to the \( k \)th equation for the \( j \)th aphorism of Category \( C_i \). These two extra subscripts are necessitated by the facts that a category encompasses many aphorisms and one aphorism may itself be governed by more than one equation.

4. Results and Discussion

The main sandhi aphorisms, their brief description (Rule), the corresponding general operator and the final, specialized equations along with the domain of operation are given below in a category-wise listing. Special notations followed are:

- The equations and conditions given as operators with three subscripts are the ones that are implementable. The ‘general operator’ specified for each aphorism typifies the aphorism’s meaning and all the conditions it becomes operative under, and provides a generalization from which the final equations are specialized. A specialized operator would thus override the ‘general operator’ with its own specialized conditions.
- The variable \( X \) denotes the sequence of letters culminating in \( x \); the variable \( Y \) denotes the sequence of letters starting with \( y \). These are used to depict special conditions that pertain to the entire word involved in the sandhi.
- Variables \( u \) and \( w \) represent the value of the letter occurring just before \( x \) and just after \( y \) respectively.
- \([ ]\) are used to club domain conditions simply in order to depict the ‘or’ condition more clearly.

4.1 \( C_1 \) Sandhi-s

1. \( \text{гуна sandhi} \)

a) \( \text{адхуна} \) || 6.1.87 ||

Rule: \( a \) or \( ō \) followed by \( i \), \( u \) (short and long) \( \rightarrow \) \( \text{гуна} \) letter \( (e, o) \) corresponding to second letter replaces both.

General operator: \( \oplus_{1,1}(x, y) = z = z_1 \) when \( x \in \{1, 2\}, y \in \{3, 4, 5, 6\} \)

\( \oplus_{1,1,1}(x, y) = z_11 = 10 \) when \( y \in \{3, 4\} \)

\( \oplus_{1,1,2}(x, y) = z_11 = 11 \) when \( y \in \{5, 6\} \)

2. \( \text{uran raparaha} \) || 1.1.51 ||

Rule: \( a \) or \( ō \) followed by \( r \) (short and long), \( l \rightarrow \text{гuna} \) letter \( (ar, al) \) corresponding to the second letter replaces both.

General operator: \( \oplus_{1,2}(x, y) = z = z_1 = z_{11}z_{12} \) when \( x \in \{1, 2\}, y \in \{7, 8, 9\} \)

\( \oplus_{1,2,1}(x, y) : z_{11} = 1, z_{12} = 17 \) when \( y \in \{7, 8\} \)

\( \oplus_{1,2,2}(x, y) : z_{11} = 1, z_{12} = 18 \) when \( y \in \{9\} \)

3. \( \text{vṛddhiresi} \) || 6.1.88 ||

Rule: \( a \) or \( ō \) followed by \( e, o, ai, au \rightarrow \text{vṛddhi} \) letter \( (ai, au) \) corresponding to second letter replaces both.

General operator: \( \oplus_{1,3}(x, y) = z = z_1 \) when \( x \in \{1, 2\}, y \in \{10, 11, 12, 13\} \)

\( \oplus_{1,3,1}(x, y) : z_1 = y + 2 \) when \( y \in \{10, 11\} \)

\( \oplus_{1,3,1}(x, y) : z_1 = y \) when \( y \in \{12, 13\} \)

4. \( \text{etvedhatyāṣṭru} \) || 6.1.89 ||

Rule: In all the following rules, vṛddhi letter \( (ai, au, ār, āl) \) corresponding to the beginning of second word, replaces both.

a. \( a \) or \( ō \) followed by the prepositions \( eti, edhati \rightarrow ai \) replaces both

b. preposition \( pra \) followed by the preposition \( esah, esya \rightarrow ai \) replaces both

c. word \( sva \) followed by \( čr \rightarrow ai \) replaces both

d. \( a \) or \( ō \) followed by the preposition \( iḥ \rightarrow au \) replaces both

e. word \( aksa \) followed by word \( īhini \rightarrow au \) replaces both

f. preposition \( pra \) followed by \( iḥ, īdḥ \rightarrow au \) replaces both

General operator: \( \oplus_{1,4}(x, y) = z = z_1 \) when \( x \in \{1, 2\} \)

\( \oplus_{1,4,1}(x, y) : z_11 = 12 \) when \( y \in \{10, 11, 12, 13\} \)

\( \oplus_{1,4,2}(x, y) : z_11 = 12 \) when \( y \in \{10, 11, 12, 13\} \)

or \( \oplus_{1,4,3}(x, y) : z_11 = 12 \) when \( y \in \{10, 11, 12, 13\} \)

\( \oplus_{1,4,4}(x, y) : z_11 = 12 \) when \( y \in \{10, 11, 12, 13\} \)

\( \oplus_{1,4,5}(x, y) : z_11 = 12 \) when \( y \in \{10, 11, 12, 13\} \)

or \( \oplus_{1,4,6}(x, y) : z_11 = 12 \) when \( y \in \{10, 11, 12, 13\} \)

5. etvedhatyāṣṭru || 6.1.89 ||

Rule: In all the following rules, vṛddhi letter \( (ai, au, ār, āl) \) corresponding to the beginning of second word, replaces both.

a. \( a \) followed by word \( ṛṭa \rightarrow ār \) replaces both

b. preposition/words \( pra, vaśara, kambala, vasana, daśa, ṛṭa \) followed by the word \( ṛṭa \rightarrow ār \) replaces both

General operator: \( \oplus_{1,5}(x, y) = z = z_{11} \) when \( x \in \{1, 2\} \)

\( \oplus_{1,5,1}(x, y) : z_{11} = 2, z_{12} = y + 10 \) when \( y \in \{7, 8, 9\} \)

or \( \oplus_{1,5,2}(x, y) : z_{11} = 2, z_{12} = y + 10 \) when \( y \in \{7, 8\} \)

or \( \oplus_{1,5,3}(x, y) : z_{11} = 2, z_{12} = y + 10 \) when \( y \in \{7, 8\} \)

or \( \oplus_{1,5,4}(x, y) : z_{11} = 2, z_{12} = y + 10 \) when \( y \in \{7, 8\} \)

6. upasargāḍti dhātus || 6.1.91 ||

Rule: \( a \) or \( ō \) at the end of prepositions followed by \( r \rightarrow \text{vṛddhi} \) letter \( ār \) replaces both. (The prepositions that qualify are: \( pra, parā, apa, ava, upa \))
General operator: $\oplus_{1,6}(x, y) = z = z_1 = z_2z_1$ when $x \in \{1, 2\}, y = 7$
$\oplus_{6,1}(x, y) = z_{11} = 2, z_{12} = y + 10$ when $X \in \{43+17+1, 43+17+2, 1+43+1, 1+16+1, 5+43+1\}$

pararūpa sandhi

7. *enī pararūpaṃ* || 6.1.94 ||
Rule: a or o at the end of a preposition followed by e or o (of a verbal root) -> second letter (e or o) replaces both.
General operator: $\oplus_{1,7}(x, y) = z = z_1 = y$ when $x \in \{1, 2\}, y \in \{10,11\}$
$\oplus_{7,1}(x, y) = z_1 = y$ when $x \in \{1, 2\}, y \in \{10,11\}, X \in \{43+17+1, 43+17+2, 1+43+1, 1+16+1, 5+43+1\}$

savānadvīgha sandhi

8. *akah savarne dirghaḥ* || 6.1.101 ||
Rule: a, i, u, r, l (short or long) followed by similar a, i, u, r, l (short or long) -> corresponding long letter replaces both.
General operator: $\oplus_{1,8}(x, y) = z = z_1 = y$ when $1 <= x <= 9, 1<= y <= 9$
All operators $\oplus_{1,8}$ are commutative.
$\oplus_{3,1}(x, y) = z_1 = y$ when $[x \in \{1, 3, 5\}, y = x+1]$ or $[x \in \{2, 4, 6\}, y = x]$
$\oplus_{3,2}(x, y) = z_1 = y + 1$ when $x \in \{1, 3, 5\}, y = x$
$\oplus_{3,3}(x, y) = z_1 = 8$ when $x, y \in \{7, 8, 9\}$

pārvarūpa sandhi

9. *enāḥ padāntādātī* || 6.1.109 ||
Rule: e or o followed by a -> first letter replaces both.
General operator: $\oplus_{1,9}(x, y) = z = z_1 = x$ when $x \in \{10, 11\}, y = 1$
$\oplus_{9,1}(x, y) = z_1 = x$ when $x \in \{10, 11\}, y = 1$

4.2 C₂ Sandhi-s

yaṇādeśa sandhi

1. *ikō yanaci* || 6.1.77 ||
Rule: i, u, r, l (short and long) followed by dissimilar vowel -> v, r, l respectively replace first letter.
General operator: $\oplus_{2,1}(x, y) = z = z_{2y}$ when $3 <= x <= 9, y <= 13$
$\oplus_{2,1}(x, y) = z_2 = 15$ when $x \in \{3, 4\}, y \in \{3, 4\}$
$\oplus_{2,2}(x, y) = z_2 = 16$ when $x \in \{5, 6\}, y \in \{5, 6\}$
$\oplus_{2,3}(x, y) = z_2 = 17$ when $x \in \{7, 8\}, y \in \{7, 8, 9\}$
$\oplus_{2,4}(x, y) = z_2 = 18$ when $x \in \{9\}, y \in \{7, 8, 9\}$

ayūya-āvāva-ādeśa sandhi

2. *ecoyavāyāvah* || 6.1.78 ||
Rule: e, o followed by āc -> ay, ayv replace the first respectively;
ai, au followed by ac -> āy, āv replace the first respectively.
General operator: $\oplus_{2,2}(x, y) = z = z_{2y} = z_{2z_2y}$ when $10 <= x <= 13, y <= 13$
$\oplus_{2,2}(x, y) : z_2 = 1, z_{2z_2} = x + y$ when $x \in \{10, 11\}, y <= 1$
$\oplus_{2,2}(x, y) : z_2 = 2, z_{2z_2} = x + 3$ when $x \in \{12, 13\}$

3. *vānto yi pratyayey* || 6.1.79 ||
Rule: a, au followed by y -> av, āv replace the first respectively.
General operator: $\oplus_{2,3}(x, y) = z = z_{2y} = z_{2z_2y}$ when $x \in \{11, 13\}, y = 15$
$\oplus_{3,1}(x, y) : z_{21} = 1, z_{22} = x + 5$ when $x = 11$
$\oplus_{3,3}(x, y) : z_{21} = 2, z_{22} = x + 3$ when $x = 13$

4. *ksayajayauk śakābyīrthā* || 6.1.81 ||
Rule: e which is the end of words kse, je, ke followed by y -> ay replaces the first.
General operator: $\oplus_{2,4}(x, y) = z = z_{2y} = z_{2z_2y}$ when $x = 10, y = 15, X \in \{42+45+10, 29+10\}$
$\oplus_{2,2}(x, y) = z_{21} = 1, z_{22} = x + 5$

avanādeśa sandhi

5. *avośi sphatīyanasya* || 6.1.123 ||
Rule: o which is the end of word go followed by a vowel -> ‘av’ replaces the first.
General operator: $\oplus_{2,5}(x, y) = z = z_{2y} = z_{2z_2y}$ when $x = 11, y <= 9, X = 11+1$
$\oplus_{5,1}(x, y) : z_{21} = 1, z_{22} = 16, z_{23} = 1$

jaśva sandhi

6. *jhalām jaśo’nte* || 8.2.39 ||
Rule: non-nasal mutes, sibilants, aspirate at the end of a word -> first letter replaced by corresponding column 3 letter.
General operator: $\oplus_{2,6}(x, y) = z = z_{2y}$ when $24 <= x <= 47, y = 0$
$\oplus_{6,1}(x, y) : z_2 = x + 5$ when $x \in \{24, 25, 26, 27, 28\}$
$\oplus_{6,2}(x, y) : z_2 = x$ when $x \in \{29, 30, 31, 32, 33, 44, 45, 47\}$
$\oplus_{6,3}(x, y) : z_2 = x – 3$ when $x = 34$
$\oplus_{6,4}(x, y) : z_2 = x – 5$ when $x \in \{35, 37, 38\}$
$\oplus_{6,5}(x, y) : z_2 = x – 7$ when $x = 36$
$\oplus_{6,6}(x, y) : z_2 = x – 8$ when $x \in \{40, 41\}$
$\oplus_{6,7}(x, y) : z_2 = x – 11$ when $x = 42$
$\oplus_{6,8}(x, y) : z_2 = x – 13$ when $x = 43$
$\oplus_{6,9}(x, y) : z_2 = x – 10$ when $x = 39$

satva sandhi

7. *samah suṣī* || 8.3.5 ||
Rule 1: word sam followed by affixes kr, kṛ, kar, kār, kur -> m of sam replaced with the combination m{s}.
General operator: $\oplus_{2,7}(x, y) = z = z_{2y} = z_{2z_2y}$ when $x = 20, y = 42, X \in \{46+1+20\}, Y \in \{42+7, 42+8, 42+1+17, 42+2+17, 42+5+17\}$
$\oplus_{7,1}(x, y) : z_{21} = 48, z_{22} = 46$

8. *pumah khayyamhāre* || 8.3.6 ||
Rule: word pum followed by column 1, column 2 which is in turn followed by a vowel, aspirate, semi-vowel or nasal -> ending m replaced with the combination m{s}.
General operator: \( \oplus_{2,8} (x, y) = z = z_2 y = z_2 z_2 y \) when \( x = 20, 34 \leq y \leq 43, 1 < w \leq 23 \)
\( \oplus_{2,8,1} (x, y) : z_{21} = 48, z_{22} = 46 \)

9. naschayaprasān || 8.3.7 ||
Rule: final \( n \) of a word except for the word praśān, followed by \( ch, th, th, c, t, t \) which is in turn followed by a vowel, aspirate, semi-vowel or nasal -> ending \( n \) replaced with the combination \( nis \).
General operator: \( \oplus_{2,9} (x, y) = z = z_2 y = z_2 z_2 y \) when \( x = 23, 36 < y < 41, 1 < w \leq 23, X \in \{43 + 1 + 44 + 2 + 23\} \)
\( \oplus_{2,9,1} (x, y) : z_{21} = 48, z_{22} = 46 \)

visarga sandhi

10. kharavasānyorvisarjanīyaḥ || 8.3.15 ||
Rule: \( r \) followed by hard consonant -> \( r \) replaced with visarga.
General operator: \( \oplus_{2,10} (x, y) = z = z_2 y \) when \( x = 17, 34 \leq y \leq 46 \)
\( \oplus_{2,10,1} (x, y) : z_2 = 49 \)

anusvāra sandhi

11. mo'nsūvāraḥ || 8.3.23 ||
mo ṛājī samah kavu || 8.3.25 ||
Rule: \( m \) followed by any consonant -> \( m \) letter replaced by \( m \) (anusvāra) (except in the case of the word sam being followed by the word rājī).
General operator: \( \oplus_{2,11} (x, y) = z = z_2 y \) when \( x = 20, 14 \leq y \leq 47, X \in \{46 + 1 + 20\}, Y \in \{17 + 2 + 40\} \)
\( \oplus_{2,11,1} (x, y) : z_2 = 48 \)

12. nasāpādāntyasā ḫahāli || 8.3.24 ||
Rule: \( n \) followed by a non-nasal mute, sibilant or aspirate (not at the end of a pada) -> \( n \) replaced by \( in \) (anusvāra).
General operator: \( \oplus_{2,12} (x, y) = z = z_2 y \) when \( x = 23, 24 \leq y \leq 47 \)
\( \oplus_{2,12,1} (x, y) : z_2 = 48 \)

13. he maparā vā || 8.3.26 ||
Rule: \( m \) followed by \( h \) which is in turn followed by \( y, l, \) or \( v \) -> the first \( m \) replaced by nasal \( y, l, v \) (i.e. ny, ml, mv) respectively.
General operator: \( \oplus_{2,13} (x, y) = z = z_2 y = z_2 z_2 y \) when \( x = 20, y = 14, w \in \{15, 16, 18\} \)
\( \oplus_{2,13,1} (x, y) : z_{21} = 48, z_{22} = w \)

14. naparā nab || 8.3.27 ||
Rule: \( m \) followed by \( h \) at the end of a pada which is in turn followed by \( n \) replaced by \( n \).
General operator: \( \oplus_{2,14} (x, y) = z = z_2 y \) when \( x = 20, y = 14, w = 23 \)
\( \oplus_{2,14,1} (x, y) : z_2 = w \)

visarga sandhi

15. visarjanīyasya saḥ || 8.3.34 ||
Rule: visarga followed by hard consonant -> visarga replaced with s.
General operator: \( \oplus_{2,15} (x, y) = z = z_2 y \) when \( x = 49, 34 \leq y \leq 46, w \in \{44, 45, 46\} \)
\( \oplus_{2,15,1} (x, y) : z_2 = 46 \)

ścūtva sandhi

16. stob ścūnāḥ ścuh || 8.4.40 ||
Rule: dentals, \( s \) followed by palatalas, \( s \) -> first replaced by its corresponding palatal, \( s \) respectively.
General operator: \( \oplus_{2,16} (x, y) = z = z_2 y \) when \( x \in \{41, 38, 33, 28, 23, 46\}, y \in \{39, 36, 29, 24, 19, 44\} \)
\( \oplus_{2,16,1} (x, y) : z_2 = x – 2 \) when \( x \in \{41, 38, 46\} \)
\( \oplus_{2,16,2} (x, y) : z_2 = x – 4 \) when \( x \in \{33, 28, 23\} \)

ścūtva sandhi

17. ścunāḥ śtuh || 8.4.41 ||
toh sī || 8.4.43 ||
Rule: [dentals, \( s \) followed by cerebra] or [\( s \) followed by \( s \) -> dentals or \( s \) replaced by cerebra] or \( s \) respectively.
General operator: \( \oplus_{2,17} (x, y) = z = z_2 y \) when \( x \in \{41, 38, 33, 28, 23, 46\}, y \in \{40, 37, 32, 27, 22, 45\} \)
\( \oplus_{2,17,1} (x, y) : z_2 = x – 1 \) when \( x = 46 \) or \( y \neq 45 \)

anunāsikā sandhi

18. yaro'nuṇāsike 'nuṇāsiko vā || 8.4.45 ||
Rule: semi-vowels \( y, v \) and \( l \) followed by nasal -> first replaced by its corresponding nasal, \( ny, nv, ml \) respectively.
General operator: \( \oplus_{2,18} (x, y) = z = z_2 y = z_2 z_2 y \) when \( x \in \{15, 16, 18\}, 19 \leq y \leq 23 \)
\( \oplus_{2,18,1} (x, y) : z_2 = 48, z_{22} = x \)

19. yaro'nuṇāsike 'nuṇāsiko vā || 8.4.45 ||
Rule: semi-vowel \( r \), mutes, sibilants followed by nasal -> first replaced by its corresponding nasal.
General operator: \( \oplus_{2,19} (x, y) = z = z_2 y \) when \( 17 \leq x \leq 46, x != 18, 19 \leq y \leq 23 \)
\( \oplus_{2,19,1} (x, y) : z_2 = x \) when \( x \in \{17, 19, 20, 21, 22, 23, 44, 45, 46\} \)
\( \oplus_{2,19,2} (x, y) : z_2 = x – 5 \) when \( x \in \{24, 25, 26, 27, 28\} \)
\( \oplus_{2,19,3} (x, y) : z_2 = x – 10 \) when \( x \in \{29, 30, 31, 32, 33\} \)
\( \oplus_{2,19,4} (x, y) : z_2 = x – 13 \) when \( x = 34 \)
\( \oplus_{2,19,5} (x, y) : z_2 = x – 15 \) when \( x \in \{35, 37, 38\} \)
\( \oplus_{2,19,6} (x, y) : z_2 = x – 17 \) when \( x = 36 \)
\( \oplus_{2,19,7} (x, y) : z_2 = x – 18 \) when \( x \in \{40, 41\} \)
\( \oplus_{2,19,8} (x, y) : z_2 = x – 20 \) when \( x = 39 \)
\( \oplus_{2,19,9} (x, y) : z_2 = x – 21 \) when \( x = 42 \)
\( \oplus_{2,19,10} (x, y) : z_2 = x – 23 \) when \( x = 43 \)

jaśva sandhi

20. jhalām jaś jhaśi || 8.4.53 ||
Rule: non-nasal mutes, sibilants, aspirate followed by soft consonants (column 3, column 4) -> first replaced by corresponding column 3 letter.
General operator: \( \oplus_{2,20} (x, y) = z = z_2 y \) when \( 24 \leq x \leq 47, 24 \leq y \leq 33 \)
\( \oplus_{2,20,1} (x, y) : z_2 = x + 5 \) when \( 24 \leq x \leq 28 \)
22. anuvārasya yai parasavarṇaḥ $|| 8.4.58 $  
Rule: anuvāra followed by semi-vowels, mutes -> anuvāra replaced by the nasal equivalent of the second.  
General operator: $\oplus_{2,22} (x, y) = z = z_2 y$ when $x = 48, 15 <= y <= 43$  
$\oplus_{2,22,1} (x, y) : z_2 = 20$ when $x \in \{16, 17\}$  
$\oplus_{2,22,2} (x, y) : z_2 = y$ when $x \in \{15, 18, 19, 20, 21, 22, 23\}$  
$\oplus_{2,22,3} (x, y) : z_2 = y - 5$ when $24 <= x <= 28$  
$\oplus_{2,22,4} (x, y) : z_2 = y - 10$ when $29 <= x <= 33$  
$\oplus_{2,22,5} (x, y) : z_2 = y - 13$ when $x = 34$  
$\oplus_{2,22,6} (x, y) : z_2 = y - 15$ when $x \in \{35, 37, 38\}$  
$\oplus_{2,22,7} (x, y) : z_2 = y - 17$ when $x = 36$  
$\oplus_{2,22,8} (x, y) : z_2 = y - 18$ when $x \in \{40, 41\}$  
$\oplus_{2,22,9} (x, y) : z_2 = y - 20$ when $x = 39$  
$\oplus_{2,22,10} (x, y) : z_2 = y - 21$ when $x = 42$  
$\oplus_{2,22,11} (x, y) : z_2 = y - 23$ when $x = 43$

23. torli $|| 8.4.60 $  
Rule 1: dentals except $n$ followed by $l$ -> dentals replaced by $l$.  
General operator: $\oplus_{2,23} (x, y) = z = z_2 y$ when $x \in \{41, 38, 33, 28\}$, $y = 18$  
$\oplus_{2,23,1} (x, y) : z_2 = y$  

24. torli $|| 8.4.60 $  
Rule 2: $n$ followed by $l$ -> $n$ replaced by nasal $l$ (i.e. $\tilde{m}$).  
General operator: $\oplus_{2,24} (x, y) = z = z_2 y = z_2; z_2 y$ when $x = 23, y = 18$  
$\oplus_{2,24,1} (x, y) : z_2 = 48, z_2 = y$  

4.3 C₃ Sandhi-s

ścūtva sandhi

1. stoh ścūnāḥ ścuh $|| 8.4.40 ||$  
śār $|| 8.4.44 $  
Rule: [palatals followed by dentals, $s$] or [ś followed by $s$] -> second replaced by palatals or ś respectively.  
General operator: $\oplus_{3,1} (x, y) = z = xz_3$ when $x \in \{39, 36, 29, 24, 19, 44\}$, $y \in \{41, 38, 33, 28, 23, 46\}$  
$\oplus_{3,1,1} (x, y) : z_3 = y - 2$ when $[y = 46]$ or [$x = 44$, $y \in \{41, 38\}$]  
$\oplus_{3,1,2} (x, y) : z_3 = y - 4$ when $x = 44, y \in \{33, 28, 23\}$

śvuta sandhi

2. śṭunāḥ śṭuh $|| 8.4.41 ||$  
na padāntāṭṭorangām $|| 8.4.42 $  
Rule: $s$ followed by dentals, $s$ -> $d$, $s$ replaced by cerebrals, $s$ respectively.  
General operator: $\oplus_{3,2} (x, y) = z = xz_3$ when $x = 45, y \in \{41, 38, 33, 28, 23, 46\}$  
$\oplus_{3,2,1} (x, y) : z_3 = y - 1$

pārvavāsvarṇa sandhi

3. jhayo ho’nyatarasyām $|| 8.4.62 $  
Rule: non-nasal mutes followed by $h$ -> $h$ replaced by the aspirate letter (column 4) corresponding to the first non-nasal mute.  
General operator: $\oplus_{3,3} (x, y) = z = xz_3$ when $24 <= x <= 43, y = 47$  
$\oplus_{3,3,1} (x, y) : z_3 = x$ when $24 <= x <= 28$  
$\oplus_{3,3,2} (x, y) : z_3 = x - 5$ when $29 <= x <= 33$  
$\oplus_{3,3,3} (x, y) : z_3 = x - 8$ when $x = 34$  
$\oplus_{3,3,4} (x, y) : z_3 = x - 10$ when $x \in \{35, 37, 38\}$  
$\oplus_{3,3,5} (x, y) : z_3 = x - 12$ when $x = 36$  
$\oplus_{3,3,6} (x, y) : z_3 = x - 13$ when $x \in \{40, 41\}$  
$\oplus_{3,3,7} (x, y) : z_3 = x - 15$ when $x = 39$  
$\oplus_{3,3,8} (x, y) : z_3 = x - 16$ when $x = 42$  
$\oplus_{3,3,9} (x, y) : z_3 = x - 18$ when $x = 43$

chātva sandhi

4. saścō’ṭi $|| 8.4.63 ||$  
Rule: non-nasal mutes followed by ś which is in turn followed by a vowel, aspirate or $r$, $r$ -> ś replaced by $ch$.  
General operator: $\oplus_{3,4} (x, y) = z = xz_3$ when $24 <= x <= 43, y = 44, 1 <= w <= 17$  
$\oplus_{3,4,1} (x, y) : z_3 = 36$

4.4 C₄ Sandhi-s

tugāgama sandhi

1. che ca $|| 6.1.73 $  
āṁmāṅsāca $|| 6.1.74 $  
dirgāṭ $|| 6.1.75 $  
padāntāṭāvā $|| 6.1.76 $  
Rule: vowel followed by $ch$ -> $t$ added.  
General operator: $\oplus_{4,1} (x, y) = z = xz_4 y$ when $x = 13, y = 36$  
$\oplus_{4,1,1} (x, y) : z_4 = 41$
2. **dhuṇāgaṇa sandhi**

   **Rule:** If a nasal (nas) is followed by a consonant, substitute dh for the nasal.

   **General operator:** \( \oplus \) (x, y) = z when x = 23, 32, y = 46

   \( \oplus \) (x, y) : z = 28

3. **tugāgaṇa sandhi**

   **Rule:** If an oṣu (tuk) is followed by a consonant, substitute θ for the vowel.

   **General operator:** \( \oplus \) (x, y) = z when x = 23, 32, y = 46

4. **nāmuṇāgaṇa sandhi**

   **Rule:** If a short vowel (ṅ) is followed by a consonant, substitute θ for the vowel.

   **General operator:** \( \oplus \) (x, y) = z when x = 23, 32, y = 46

5. **śuṇāgaṇa sandhi**

   **Rule:** If a long vowel (ṅ) is followed by a consonant, substitute θ for the vowel.

   **General operator:** \( \oplus \) (x, y) = z when x = 23, 32, y = 46

6. **śvādi sandhi**

   **1. etattadōh sulopo ‘koranaśamāse halī**

   **Rule:** If a consonant (n) is followed by a vowel, substitute a consonant.

   **General operator:** \( \oplus \) (x, y) = z when x = 23, 32, y = 46

   \( \oplus \) (x, y) : z = 28

7. **so ‘ci lope cetpādaṁpāramān**

   **Rule:** If a consonant (n) is followed by a vowel, substitute a consonant.

   **General operator:** \( \oplus \) (x, y) = z when x = 23, 32, y = 46

   \( \oplus \) (x, y) : z = 28

8. **lopah śākalyasya**

   **Rule:** If a consonant (n) is followed by a vowel, substitute a consonant.

   **General operator:** \( \oplus \) (x, y) = z when x = 23, 32, y = 46

   \( \oplus \) (x, y) : z = 28

9. **oto gāryasya**

   **Rule:** If a consonant (n) is followed by a vowel, substitute a consonant.

   **General operator:** \( \oplus \) (x, y) = z when x = 23, 32, y = 46

   \( \oplus \) (x, y) : z = 28

10. **hali sarveṣām**

    **Rule:** If a consonant (n) is followed by a vowel, substitute a consonant.

    **General operator:** \( \oplus \) (x, y) = z when x = 23, 32, y = 46

    \( \oplus \) (x, y) : z = 28

5. **Conclusions**

   In spite of there being almost 2500 individual letter-level rules (Table 2), this new schema that directly maps the patterning in the Pāṇiniian aphorisms in a simple and effective way, which ensures that we arrive at a total of just 110 equations. Clearly, this is a computationally lean way of calculating the result of sandhi operations. The results represent a computational model to process a majority of the euphonious conjunctions in Sanskrit. The work also demonstrates the simplicity with which euphonic conjunctions can be handled by adopting Pāṇini’s precise scheme for rule representation.

   A main strength of this modeling approach is that it is deterministic, as against the probabilistic methods adopted till now for sandhi operations. Determinism is inherent in Pāṇini’s sandhi rules, which indeed specify how sandhi-s are formed and not how they are broken up, and this determinism has been uniquely tapped and modeled in this work. Traditional AI methods such as hidden Markov models, which have hitherto been applied for Sanskrit processing [3-7], assume relevance in the sandhi-splitting approach in which there are inherent ambiguities, rather than in the sandhi-building approach which is modeled here.

   The five main operators and all the 110 derived equations designed and presented in this work, form the immediate basis for directly realizing crucial applications of sandhi-processing such as subtext searching.

References

[1] Briggs Rick, “Knowledge Representation in Sanskrit and Artificial Intelligence”, RIACS, NASA Ames Research Center, AI Magazine, 1985.

[2] Dikṣīta Bhaṭṭoḍi, Siddhānta-kaumudi, Translated by Śrīśa Candra Vasu, Volume 1, Motilal Banarsidas Publishers, Delhi, 1962.

[3] Huet Gérard, “Automata Mista”, Festschrift in Honor of Zohar Manna for his 64th anniversary, Taormina,
Sicily, July 2003, in Verification: Theory and Practice: Essays Dedicated to Zohar Manna on the Occasion of His 64th Birthday, Ed. Nachum Dershowitz, Springer-Verlag LNCS vol. 2772, 2004, pp. 359-372.

[4] Huet Gérard, “Lexicon-directed Segmentation and Tagging of Sanskrit”, in XIth World Sanskrit Conference, Helsinki, Finland, Aug. 2003. Final version in Themes and Tasks in Old and Middle Indo-Aryan Linguistics, Eds. Bertil Tikkanen & Heinrich Hettrich. Motilal Banarsidass, Delhi, 2006, pp. 307-325.

[5] Huet Gérard, “Shallow Syntax Analysis in Sanskrit Guided by Semantic Nets Constraints”, Proceedings of International Workshop on Research Issues in Digital Libraries, Kolkata, Dec. 2006. Editors Prasenjit Majumder, Mandar Mitra and Swapan K. Parui, ACM Digital Library.

[6] Hellwig Oliver, “SanskritTagger, a Stochastic Lexical and POS Tagger for Sanskrit”, Peer-reviewed Proceedings of the First International Sanskrit Computational Linguistics Symposium, France, 2007.

[7] Hyman Malcolm D., “From Paninian Sandhi to Finite State Calculus”, Sanskrit Computational Linguistics: First and Second International Symposia, Revised Selected and Invited Papers, ISBN:978-3-642-00154-3, Springer-Verlag, 2009.

[8] Rao T. R. N., Kak Subhash, Computing Science in Ancient India, Center for Advanced Computer Studies, University of Southwestern Louisiana, 1998.

[9] Rao T. R. N., Kak Subhash, The Panini-Backus Form in Syntax of Formal Languages, Center for Advanced Computer Studies, University of Southwestern Louisiana, 1998.

[10] Varadarāja, Laghu-siddhānta-kaumudī, Translated with commentary by Pandit Viśvanātha Śāstri Prabhākara, Motilal Banarsidas Publishers, Delhi, 1989.

Websites

[11] Göttingen Register of Electronic Texts in Indian Languages (GRETIL), www.sub.uni-goettingen.de/ebene_1/fiindolo/gretil.htm.

[12] Sanskrit Heritage site, Program and documentation, Gérard Huet, pauillac.inria.fr/~huet/SKT/.

Rama N. B.Sc. (Mathematics), Master of Computer Applications, Ph.D. (Computer Science). Employment: Anna Adarsh College, Chennai; Bharathiyar Women’s College, Chennai; Presidency College, Chennai, India. Has 20 years of teaching experience including 10 years for PG, has guided 15 M.Phil. students; Chairperson of the Board of Studies in Computer Science for UG, and Member, Board of Studies in Computer Science for PG and Research at the University of Madras. Current research interest: Program Security. Member of Editorial cum Advisory Board of Oriental Journal of Computer Science and Technology.

Meenakshi Lakshmanan. B.Sc. (Mathematics), Master of Computer Applications, M.Phil. (Computer Science), currently pursuing Ph.D. (Computer Science) at Mother Teresa Women’s University, Kodaikanal, India and pursuing Level 4 Sanskrit (Samartha) of the Sanskrita Bhāṣā Pracārini Sabha, Chittoor, India. Employment: SRA Systems Pvt. Ltd., currently Head of the Department, Department of Computer Science, Meenakshi College for Women, Chennai, India. Is a professional member of the ACM and IEEE.