Computational Referencing System for Sanskrit Grammar

Baldev Ram Khandoliyan, Ram Kishor
SSIS, Jawaharlal Nehru University,
New Delhi
{Baldevramjnu, crks113}@gmail.com

Abstract:
The goal of this project was to reconstitute and storage the text of Āṣṭādhyāyī (AD) in a computer text system so that everyone may read it. The proposed work was to do study the structure of AD and to create a relational database system for storing and interacting with AD. The system is available online, including Devanāgari Unicode and other major Indian characters as input and output. MS SQL Server, a Relational Database Management System (RDBMS)-based system, and Java Server Pages (JSP) were used. For AD, the system works as a multi-dimensional interactive knowledge-based computer system. The approach can also be applied to all Sanskrit sūtra texts that have a similar format. Sanskrit heritage texts are projected to benefit from the system's preservation and promotion. A research is being made here for preparing an AD text as a computer aided dynamic search, learning and instruction system in the Indian context.

Keywords: Computational Linguistics, Sanskrit language, Sanskrit corpus, Sanskrit computational lexicography, Computer Text System, Āṣṭādhyāyī, Sanskrit grammar.

1. Introduction
Computational Linguistics (CL) does research and development to make Human Computer Intelligent Interaction (HCII) having language as the means of communication. Computational Linguists worldwide are taking great interest in Sanskrit language, texts and particularly in Sanskrit grammar. With the help of computer tasks like electronic storage, access and preservation of Sanskrit corpus, Sanskrit computational lexicography, search engines, indices and glossary etc. are going to made. A computer system for Sanskrit grammar based on the text of AD has developed to store and make accessible it in a interactive way with user friendly format. In the present time five kinds of CL research work for Sanskrit based on is in process. These are the following types:

1. analysis and description of Sanskrit structure;
2. developing learning/teaching systems within the Paninian frame;
3. inferring from AD;
4. its study as a primary model of knowledge (semiotics).
5. computerizing and storing the text;

The last one area may be modified into the study and storing of AD as a text as a Śāstra. A full access to the text will be made possible by reconstituting or organizing it into a computer text system. Computer system for AD is essential for search and e-learning/teaching of Sanskrit grammar. A tentative proposal is being made here for the reconstruction of AD text into database for a computer aided learning and instruction system in the Indian context. The reconstruction of AD into database is required for the development of the proposed system. The Reconstruction will consist, for each sūtra, of

(i) Analysis into pada
(ii) Restoring adhikāra-anuvṛtti elements
(iii) Identifying the samāśa, if required
(iv) Marking the vibhaktis of each pada
(v) Re-ordering the pada in Case(Kāraka)-orders as required-

| Case | Symbol |
|------|--------|
| 5-7-6-1 | Kṛtā marks the substitute. |
| 5-7-1 | Sambandha marks the thing that is removed due to substitution. |
| 6-1-7 | Apādāna marks the thing that appears before the substitution. |
|    | Adhikaraṇa marks the thing that appears after the substitution. |
(vi) Translating the sūtras into Hindi, English.

1 Kapoor, Kapil, Dimensions of Pāṇini Grammar, p. 125
Support texts will be separately included: Dhātuvāpaṭha (DP), Gaṇapāṭha (GP), Pratyāhārasūtras (PS), Lingāmucāśana (LS) and Unādisūtras (US). Also IT-Saṁjñā rules and Pāṇinīya Śikṣā as well in original Sanskrit, with provision for recall and display along with the pertinent sūtra.

For reference, indices have been organized as follows:

1. Alphabetical list of sūtras
2. Thematic grouping of sūtras [65 themes according to Siddhānta Kaumudī (SK)].
3. Pratyāhāra list and Pratyāhāra-generation programme
4. IT-Saṁjñās dictionary
5. Dictionary of affixes
6. Dictionary of technical terms
7. Typical declension paradigms (83 paradigms)
8. Typical conjugation paradigms (23 paradigms)
9. Typical Examples of each of the seven kinds of siddhi
10. Sandhi-enumeration and corresponding rules
11. Concordance of pada in AD
12. Enumeration of representative, frequent nouns and verbs for declension and conjugation paradigms.

Further, the power of this system will increase enormously if explanatory comments in major Tīkas on different AD rules are made available as reference system for each rule that has been commented upon.

2. Structure and Organization of AD

AD has 8 chapters divided into 4 padsas. A sūtra or rule is referenced as x.x.x (x adhyāya, x pāda, x sūtra). For example sūtra 1.1.1 (vṛddhirādāic) is adhyāya one, pāda one and sūtra one.

The components of AD are as follows:

1. Phonetic component
   1. Phonems (aksarasamāmāṇyā - 14 sūtras called śiva-sūtras) (AS)

2. Pratāhāras (sigla)
3. Rule base (sūtrapāṭha - 4000 sūtras - 3983 in kāśikāvṛtti) (SP)
   1. List
      1. Sound (varna samuccaya)
      2. Affixes – SuP, Tiṅ, Kṛt, Taddhīta, Saṅ, Strī
   2. gaṇapāṭha (other pertinent items like primitive nominal bases, avyayas) (GP)
   3. Lingāmucāśana (LS)

The AS, DP, and the GP can be called the three most basic databases of the Pāṇinian system containing duly arranged and structured data. The SP is Pāṇini’s comprehensive rule base for Sanskrit.

3. The reconstruction of AD

The reconstruction of AD into database is required for the development of the proposed system. The Reconstruction will consist, for each sūtra, of

   (vii) Analysis into pada
   (viii) Restoring adhikāra-anuvṛtti elements
   (ix) Identifying the samāśa, if required
   (x) Marking the vibhaktis of each pada
   (xi) Re-ordering the pada in
      5-7-6-1
      6-1-7
      Case-orders as required.
   (xii) Translating the sūtras into Hindi, English.

Support texts will be separately included: Dhātuvāpaṭha (DP), Gaṇapāṭha (GP), Pratyāhārasūtras (PS), Lingāmucāśana (LS) and Unādisūtras (US). Also IT-Saṁjñā rules and Pāṇinīya Śikṣā as well in original Sanskrit, with provision for recall and display along with the pertinent sūtra.

For reference, indices will be organized as follows:

13. Alphabetical list of sūtras
14. Thematic grouping of sūtras [65 themes according to Siddhānta Kaumudī (SK)].
15. Pratyāhāra list and Pratyāhāra-generation programme

---

2 Ibid. p. 125
3 Based on “The system of Pāṇini” (Language in India, volume 4:2 February 2004) at http://www.languageinindia.com/feb2004/panini.html
4 Kapoor, Kapil, Dimensions of Pāṇini Grammar, p. 125
16. *IT-Saṃjñas* dictionary
17. Dictionary of affixes
18. Dictionary of technical terms
19. Typical declension paradigms (83 paradigms)
20. Typical conjugation paradigms (23 paradigms)
21. Typical Examples of each of the seven kinds of *siddhi*
22. *Sandhi*-enumeration and corresponding rules
23. Concordance of *pada* in AD
24. Enumeration of representative, frequent nouns and verbs for declension and conjugation paradigms.

Further, the power of this system will increase enormously if explanatory comments in major *Tikas* on different AD rules are made available as reference system for each rule that has been commented upon.

### 4. Research Methodology:

For the development of a computer system of AD the methodology will be used as follows:

- To study the structure of AD
- To study the structure of lexicons (DP, GP, LS)
- To study the structure of database
- To create a database for AD system which would consist of:
  1. Explicit reconstitution of the AD-*sūtra*;
  2. *pada*-artha;
  3. *tattvārtha* through elucidation – based on *vārttika* and *Mahābhāṣya* wherever needed and available. This will be presented in two levels (a) simplified for average reader, and (b) scholarly for advanced students and scholars. Further –
  4. *sādhāraṇa*-bhāṣā-*ṭikā*;
  5. *uddhāraṇa*, *pratyuddhāraṇa*;
  6. *vākya*-prayoga,
  7. reference to further portions of texts, if necessary.
- To develop the necessary front end and search program.

The methodology of comparative study and analysis used in Sanskrit based Natural Language Processing (NLP), and techniques of software engineering will be also used for this work. The authentic edition of AD and supported texts mention above will be used.

### 5. Development of the *Aṣṭādhyāyī* System:

A dynamic web application cum-indexer has developed under this research. This web application is developed in the front-end of Apache Tomcat Web server using JSP and Java servlets. And its data is in Unicode data files along with RDBMS in MS SQL server. The MS-JDBC connection is used to link the front-end to the database server. The system is available online at [http://sanskrit.jnu.ac.in](http://sanskrit.jnu.ac.in) with input and output in Devanāgarī Unicode and in other major Indian scripts. The following model describes the multi-tiered architecture of the *Aṣṭ* system is given below (Fig. 1).

![Fig. 1: Program architecture](http://sanskrit.jnu.ac.in)

### 5.1. Process flow of the system:

There are various ways to give input to the system e.g. Direct Search, Alphabet search and search by the structure of the text, tree search in Devenāgarī Unicode and major 10 Indian scripts (*Punjābī, Assamese, Bengāli, Oriyā, Telugu, Tamil, Kannada, Malayālam, Maharāṭhī, Gujarāṭī*).
Step I: Preprocessing.

Preprocessing a word mainly consists of transformation of a raw data required to facilitate further processing. For example – processor can remove any non Devanāgarī and other Indian scripts characters, punctuations that may have been inadvertently introduced by the user like “#” in AD.

Step II: AD Search and Database.

At this step, the system can make an indexed list of exact and partially matching words. Getting the query as an input, the system, after a light preprocessing, sends it to the database. If the word/number has its occurrence in the database, the system is giving the output.

Step III: Output level-1.

At this stage, the system is giving all the occurrences of the searched query with its numerical reference in a hyperlinked mode.

Step IV: Output level-2.

Clicking on hyperlinked sūtra/word/number, system shows its original place and its full reference in the text of AD with sūtra viccheda marking the vibhaktis of each pada and adhikāra-anuvṛtti elements of respective sūtra. It also asks for further information from supported texts (DP, GP, PS, LS and US) and also having link with lists of indices (like thematic grouping of sūtras, pratyāhāra list, pratyāhāra-generation programme, IT-Saṁjās dictionary, dictionary of affixes, dictionary of technical terms, typical declension paradigms, typical conjugation paradigms, typical examples of each of the seven kinds of siddhi, Sandhi-enumeration and corresponding rules, concordance of pada in AD, enumeration of representative, frequent nouns and verbs for declension and conjugation paradigms and supplementary texts etc.) A hyperlink also is for IT-Saṁjā rules and Pāṇinīya Śikṣā as well in original Sanskrit, with provision for recall and display along with the pertinent sūtra.

The Kāśikā is a joint work of Jayāditya and Vāmana. The Kāśikā is a running commentary on Pāṇi’s Aṣṭādhyāyī and its merit consists in the lucid manner in which it has explained the sūtras of Pāṇini, clearly indicating all the Anuvṛttis and giving numerous illustrations for each rule. Sometimes the Kāśikā gives us information which we could not possibly have obtained from any other sources.

The Hindi English translation of sūtra meaning The Hindi meaning of Sūtras is translated by student. The explanation of Sūtra is given. Sūtra Padchheda, Sūtra Vibhakti, Sūtra Sandhi, Sūtra Samasa, Anuvṛti and Sūtra Sanskrit Vṛtti with Examples. Hindi meaning of Sūtra with detailed information also given. The English meaning is student. For Other Indian Languages translation system is used like Google translation, Bing Translation etc.

Step V: Output- final level.

Here, the system gives a list of online tools like e-learning model, TTS for reading sūtra, vṛtti etc, and also have the facility to do morphological analysis of the query with the help of POS tagger and subanta tihant, sandhi, kṛdanta analyzers.

5.2. Front-End of the AD System:

The front-end of the system is developed in UTF-8 enabled Java Server Pages (JSP) and HTML. The front-end of the software enables the user to interact with the computer system of AD with the help of

http://sanskrit.jnu.ac.in/post/post.jsp
Available at http://sanskrit.jnu.ac.in
Apache Tomcat web-server. The JSP technology helps to create web based applications combining Java code and displays the results as HTML. The web server runs the Java code and displays the results as HTML. For this system, there are two pages, one is the main search page and the other is cross-referential search/connect page which connects the searched query in different online e-learning, TTS and linguistic resources.

In the proposed program there are two layer search facility. First, the string entered in text-box will search in table1 which is available in the row of sūtras. These sūtras can be listed with Pāṇini’s sūtra number. After clicking on desired sūtras at display page, then detail about that sūtra displays in the following form:

AD sūtra number and SK sūtra number, sūtra with samādhi and with samādhi-viccheda, vṛttī with and without samādhi, anuvṛttā pada, anuvṛttā sūtra, adhikāra sūtra on searched sūtra, name of prakaraṇa of SK, technical term used in sūtra etc.

After this if user wish to know the meaning of technical term of sūtra or to know about adhikāra sūtra or detail of anuvṛttā sūtras than user have facility to click on desired technical term. The meaning of searched technical term displays after clicking hyperlinked term. To provide the meaning of technical term can be second level search which is from table2.

The work is connected all over world through internet after completing it so at the same time users from all over world can use the system. The program is developed on sever based for fast searching and getting the result. To damage the information of the system will not be possible as the information will be stored in database server. Here the storage and display of information is in Unicode so Font problems are resolved.

5.3. The Back-End of the System:
The back-end is built in two RDBMS tables that include co-relative data tables. Through JDBC connectivity, this Tomcat server-based programme connects to MS-SQL Server 2005 RDBMS. In the first table there is first level information with following column:

sutra_id ; sūtra_Pāṇini ; sūtra ; sūtra_kauṃudi ; sūtra_type ; sūtra_sandhi ; anuvṛtta_pada ; anuvṛtta_sūtra ; sūtra_vṛtti ; vṛtti_sandhi ; vṛtti_tech_words ; sūtra_adhikara ; kauṃudi_prakaraṇa.

In the second table there is the explanation of technical term which displays after clicking the technical term of sūtra. In this table there are three columns:

tech_id ; tech_word and explaination.

6. The Snapshots of the AD Indexer

Snapshot 6.1.

Snapshot 6.2.

Snapshot 6.3.
Besides creating database layouts, we also converted the data into forms that can be displayed on the web at Sanskrit.jnu.ac.in/AD the continuous text and padapātha are displayed in our general text display and additional data is displayed in the kramapatha web-based reader and index interface. The reader displays the sūtra in Devanāgarī script, followed by the words with sandhi analyzed (padapātha). Information in the comment field and several other fields from the sūtra and pada files is displayed below these. Two additional rows are reserved for a presentation of a comprehensive paraphrase of the sūtra and translation, yet to be composed. Clicking a word in the padapātha displays its inflectional information, stem, and lexical tag in a box beneath the word. A menu displays the possible commands, one of which allows configuring the display to suit one's preferences. Another opens a table of contents that allows one to navigate the text easily. The dynamic index to the reader allows focused access to various sorts of information. An alphabetic list of terms is displayed at the left. Selecting an item in the list displays the sūtras in which the term occurs in the tall box in the middle sary of the stem in the box at the lower right.

References:

1. Jijnasu, Pt. Brahmadatt, (2003). Aṣṭādhyāyi Bhāṣya Vṛtti, Ramlal Kapoor Trust, Haryana.
2. Ketre, Sumitra M. (1989) Aṣṭādhyāyi of Pāṇini, Motilal Banarasidas, Delhi.
3. Dīkṣita, Dr. Pushpa (2007). Aṣṭādhyāyi Sahajabodha, vol. 1–4, Pratibha Prakashan, Delhi.
4. Kumar, Prof. Avnindra (1996) Aṣṭādhyāyi-padānukramkośa, Parimal Publications, Delhi.
5. Dixit, Dr. Pushpa (2010) Aṣṭādhyāyiśāntrapāth, Samskrit Bharati, New Delhi.
6. Krishnacharya, T.R. (2005) Brhadādhūtārūpāvaliḥ, Sri Sri Jagadguru Shankaracharya Maha Samsthānam Dakshinamnaya Sri Sharada Peetham, Sringeri.
7. Chandra Pt, Īśvara (2010) Kāśikārītīḥ, Chaukhamba Sanskrit Pratishthan, Delhi.
8. Mahābhāṣyam, vol 1–7, Mimasaka, Yudhishthiro and Sudhyumanacharya, Ramlal Karpoor Trust, Bhivani, 2010
9. Jigyasu, Pt.Brahmadatta (ed.) (1998) Panini-Astadhayi, Ramlal Kapoor trust, Sonipat.
10. Vasisht, Jagdish Prasad (2013) Pāṇiniyaśīkṣā, Vandana Smriti Prakashna, Bhivani.
11. Vasudev Lakshman Shastri Panishikar (ed.) (2002) Siddhāntakaumudī, Tattvabodhinīvyākhyā, Chaukhamba Sanskrit Pratishtan, Delhi.

7. Web Availability

Besides creating database layouts, we also converted the data into forms that can be displayed on the web at Sanskrit.jnu.ac.in/AD the continuous text and padapātha are displayed in our general text display and additional data is displayed in the kramapatha web-based reader and index interface. The reader displays the sūtra in Devanāgarī script, followed by the words with sandhi an alayed (padapātha). Information in the comment field and several other fields from the sūtra and pada files is displayed below these. Two additional rows are reserved for a presentation of a comprehensive paraphrase of the sūtra and translation, yet to be composed. Clicking a word in the padapātha displays its inflectional information, stem, and lexical tag in a box beneath the word. A menu displays the possible commands, one of which allows configuring the display to suit one's preferences. Another opens a table of contents that allows one to navigate the text easily. The dynamic index to the reader allows focused access to various sorts of information. An alphabetic list of terms is displayed at the left. Selecting an item in the list displays the sūtras in which the term occurs in the tall box in the middle sary of the stem in the box at the lower right.

References:

1. Jijnasu, Pt. Brahmadatt, (2003). Aṣṭādhyāyi Bhāṣya Vṛtti, Ramlal Kapoor Trust, Haryana.
2. Ketre, Sumitra M. (1989) Aṣṭādhyāyi of Pāṇini, Motilal Banarasidas, Delhi.
3. Dīkṣita, Dr. Pushpa (2007). Aṣṭādhyāyi Sahajabodha, vol. 1–4, Pratibha Prakashan, Delhi.
4. Kumar, Prof. Avnindra (1996) Aṣṭādhyāyi-padānukramkośa, Parimal Publications, Delhi.
5. Dixit, Dr. Pushpa (2010) Aṣṭādhyāyiśāntrapāth, Samskrit Bharati, New Delhi.
6. Krishnacharya, T.R. (2005) Brhadādhūtārūpāvaliḥ, Sri Sri Jagadguru Shankaracharya Maha Samsthānam Dakshinamnaya Sri Sharada Peetham, Sringeri.
7. Chandra Pt, Īśvara (2010) Kāśikārītīḥ, Chaukhamba Sanskrit Pratishthan, Delhi.
8. Mahābhāṣyam, vol 1–7, Mimasaka, Yudhishthiro and Sudhyumanacharya, Ramlal Karpoor Trust, Bhivani, 2010
9. Jigyasu, Pt.Brahmadatta (ed.) (1998) Panini-Astadhayi, Ramlal Kapoor trust, Sonipat.
10. Vasisht, Jagdish Prasad (2013) Pāṇiniyaśīkṣā, Vandana Smriti Prakashna, Bhivani.
11. Vasudev Lakshman Shastri Panishikar (ed.) (2002) Siddhāntakaumudī, Tattvabodhinīvyākhyā, Chaukhamba Sanskrit Pratishtan, Delhi.
12. Sharma, Rama Nath (2003) The Aṣṭādhyāyi of Pāṇini, Munshiram Manoharlal Publishers Pvt. Ltd., Delhi.
13. Joshi, S.D. and Roodbergen J.A.F. (2007) The Aṣṭādhyāyi of Pāṇini, vol 1-12, Sahitya Akedemi.
14. Vasu Śriśā Chandra (2003). The Aṣṭādhyāyi of Pāṇini, vol. I-II, Mitalal Banarashidas, Delhi.
15. Mimasaka, Yudhishthir (2010). Uṇādi-Kośa, Ramlal Karpoor Trust, Bhivani.
16. Chaturvedi Giridharsharma (ed.) (2004). Vaiyakarasiddhāntakaumudī with Bālamanorama and Tattvabodhini Tikā, Motilal Banarasidass, Delhi.
17. Pandey, Gopal Dutt (2007)Vaiyakarasiddhāntakaumudī with Bālamanorama Tikā, Chaukhmba Surbharati Prakashana, Varanashi.
18. Shastri, Guru Prasad (ed) (1999). Vyakaraṇa-Mahābhāṣya with Pradipoddhat Tikā, Pratibha Prakashan, Delhi.