8.18 Lexicographic Traditions in India and Sanskrit

Malhar Kulkarni, IIT, Mumbai

Abstract—Lexicographic Tradition of Sanskrit (Vedic age, Classical age [Amarakosa, other 11 lexica, analysis, the concept of synonymy and homonymy], Modern age [Wörterbuch, Monier Williams, Encyclopaedic Dictionary of Sanskrit on historical principles]... Grammatical lexicon... Application of Panini's grammar to create new lexica)

Index Terms—Dictionary, Lexicographic, Lexica, Panini, Sanskrit, Synonymy, Homonymy

I. INTRODUCTION

This paper is aimed at two things: 1) to take an overview of the lexicographic tradition of Sanskrit, highlighting the main features, and 2) to present a model of Sanskrit grammar towards developing new lexica of newly coined terms for the purpose of creation and use of scientific and technological terminologies. The lexicographic tradition of Sanskrit is divided under three heads - a) Descriptive, b) Historical and c) Etymological. We here, propose to give a brief overview of the first two classes. Amongst them, we can demarcate three stages, chronologically-the Vedic age, the classical age and the modern age.

II. VEDIC AGE

Preservation of Vedic literature was the biggest motivation for the Linguistic activity in India some approximately 2500 years ago. This activity resulted in segmenting Vedic sentences into words, and words into root-suffix components. In this activity were rooted morphological, phonological as well as morphophonological theories that were developed and applied by Panini and his tradition in the grammar of Sanskrit. The oldest known lexica of Sanskrit are entitled Nighantu (700BC) on which Yaska is believed to have written a commentary called Nirukta. There are believed to be several works known as Nighantu. These lexica are domain specific. They have only Vedic words as lexemes. We find references in the Pali literature to Nighantu as well. The Nighantu have arranged lexical material from the point of view of Synonymy as well as Homonymy. Yaska in his commentary tried to explain these word-forms from the etymological point of view.

III. CLASSICAL AGE

Researches in the field of Philology and Linguistics in the past two hundred years or so have made available to us many lexica in the printed form. We propose to take stock of some of these in the present section and to comment on their methodologies.

A. Amarakosa

The first and the foremost popular name of a lexicon in Sanskrit is that of Amarakosa written by Amarasimha before 6th century AD. Much has been said about this lexicon and its arrangement. In fact, in one of the most valuable modern contributions to the subject at hand, namely, “History of Sanskrit Lexicography” by M. M. Patkar (being the revised version of the Ph.D. dissertation submitted to the University of Mumbai), the author devotes entries 9-27 out of the total 106 to Amarakosa and the commentaries thereupon and the mention of nearly 40 commentaries on this work in the Catalogues Catalogorum, speaks volumes of the importance and the popularity of this lexicon.

B. Arrangement of Amarakosa

The name of this lexicon is “Nāmaṅgaṅ-ānuśāsana”, which means a work that deals with the lexemes and their genders. It is also known as Trikāndaṭasa and as the name suggests, it is divided into three chapters. Each chapter is further divided into sub sections called vārgas. In the 1st Chapter, also known as Svargadākāna, there are nine subsections—Svarga (heaven), vyo ma din (Ether and space), kāla (time), dhū (intellect), śaśa (mountains), nāṭya (dramaturgy), pātalabhogi (nether world), nāraka (hell), and vāri (water). In the second chapter, known as bhumyaṅdikāna, there are ten subsections— Bhūmi (earth), pura (city), śaśa (mountains), vanavasadi (Forest herbs), simhādi (animal), ṛ (human), brahma (Brahmin), kṣatriya (warrior), vaisya (trader) and śudra. The last chapter, known as sāmānyakāṇḍa, has five subsections— Viṃkṣaṇīghnaṃvarga (adjectives), samkīra (miscellaneous), nārtha (homonyms), avyaya (indeclinables), līṅgaḍisamgraha (miscellaneous).

Apart from a small subsection of the third chapter, this entire lexicon is devoted to Synonymy. In this small subsection the words are arranged after the final consonants. The formal arrangement is the unique aspect of the arrangement of this lexicon. The gender of a particular word is demonstrated by actually using the word in that gender. In the introduction, the author has clarified the point. He says—

1) prāyaśca rūpabhedaṃ sāhacaryācācācā cākaṭācāt / stripuṃnapaṃsakaṃ jāeṣaṃ tadviṣeṣāvidheḥ kva ca iti //

(Normally, the gender of a word is distinct from the difference of the form. Sometimes that is to be ascertained from the association of the other synonymous words, sometimes however, it is to be ascertained from the word which it qualifies.) Thus for example—

2) kaumodakī gāḍā khaḍgo nandakah kauṣṭubha maṇīḥ /

(The formal difference between “gāḍā” and “khaḍga” tells the gender difference between two
words. In the same line the gender of the words “nandakah”, etc, is ascertainable from the association of it with the other words in the line. There is a line at the end of dhivarga of the first chapter-

(3) guñe sukładāyah puṣi guṇālīgās tu tadvati/
(The words “sukla”, etc which denote the quality, are in masculine. However, when they denote the qualified, they take the gender of the qualified.

(4) trilīgāṃ triśviti padaṃ mithune tu dvayoritī/ 
niṣiddhaliṃgaṃ sēgarthāṃ tvaṁtathāda na pūrvabhāk/ 
(The word “trīṣ” [in all the three genders] is mentioned after the word which appears in all the three genders, the word “dvayaḥ” [in two genders] is added after the words which appear in masculine as well as in feminine. The negation of a particular gender indicates that that word appears in the rest of the genders. The words, “tu”, “anta” and “atha” added before, suggest the demarcation of the previous synonym. Thus-

(5) vyomayāṇam vimāno 'strī 
This indicates that the words vimāna and vyomayaṇa are not to be used in feminine. The instruction is not extended to other two genders.

(6) tīkto 'maśca rasāḥ puṃsi tadvatsu śadāmi 
trīṣu/ 
The words tīkta, amla, etc. are used in masculine gender when used to indicate the fluids. However, when used to indicate the entity possessing them they can be used in all the three genders. This is indicated by the word trīṣu used in the line.

C. Features of Amarakośa

Following are some of the features of this lexicon:
- Arranged in the verse form.
- No alphabetical arrangement.
- Classification of words according to Synonymy.
- Classification of words hyponymy. Names of Gods are stated first and then synonyms of each subset of god are stated. It records ISA relations.
- Syntactic features of words are pointed out. (Ex-adjective relation mentioned).
- The important feature is that words are recorded not in the form of stem but in the form of inflected forms. That does not deny the awareness of the type of the stem on the part of the Lexicographer. Some words are mentioned in big compounds.

(7) śukla-śubhra-śuci-śveta-viśada-śyeta-paṇḍarāḥ
Only stems are mentioned here. Thus it is noted that the processes of grammatical alteration of sounds, augmentation etc. are known to Amarasimha. Apart from Amarakośa several lexicis are constructed; many of which are often quoted by traditional commentators. Some of them are–

Vajrayanti of Yādavaparākṣa
Halayudha of Halayudhabhātta
Abhidhānacintāmaṇi of Hemacandra

D. Other Lexica

These lexicis were constructed more or less in the style of Amarakośa. There were some other lexica available in this period. Let us look at some of them closely. They are—
1. Namamālīka of Bhoja-11th Century
2. Siddhāśabdarṭava of Sahajakirti-17th Century
3. śāradīyakhyānamālī of Harsakirti-17th Century
4. Paryāyasaṃgātakāna of Dhananjaya Bhātta
5. Kṣākalphatru
6. Nārāṭharaṇamālī of Irugapya Dandadhinatha-14th Century
7. Nānārāmanjari of Raghava
8. Dhāraṇikosa of Dharnadasa-12th Century
9. sivakosa of SivadattaMisra
10. Ekārthanaṃmaṇā-īvyakṣaranāmaṇā of Saubhari
11. Paramānandiyavarnamāla of Makrandaṇa

They can be classified broadly into three groups

(i) Lexicon of Synonyms- 1-4
(ii) Lexicon of Homonyms- 5-10
(iii) Lexicon of both Synonyms and Homonyms- 11

E. Arrangement of These Lexica

It is worthwhile to note the system of arrangement as well as some of the features of some of these lexicis. The lexica of Synonyms are normally arranged topic wise. Lexica of Homonyms are constructed in a somewhat different manner. Thus for example, in the “ś” mentioned above, words are arranged after the final consonants and the number of syllables. The homonyms are as a system arranged in the quarter as well as in half verse or sometimes in the entire verse. In this the conjunct “kṣ” in the alphabet comes after h.

Thus if the words ending is ‘k’, we have following verses-

(8) śūko yaśaśi padye syāt lokas tu bhuvane jānē 
(The word ‘śūko’ is used in the sense of fame as well as a verse; the word ‘śūka’, however, is used in the sense of world and people.) In this, it is noted that one word is treated only in a quarter of the verse. Amongst such cases words having only two
vowels, namely, śloka, loka, stoka, etc. appear first. Then come the words- ānaka, sāyaka and jambuka, etc. with three vowels, and then the words with four vowels- uktakā, etc. The same procedure is followed everywhere.

Later on we find a verse half portion of which deals with one word. Thus-

(9) śulkaṃ ghaṭe vivāhārthe jāmātūr ghyate ca yat/

(The word ‘śulka’ is used in the sense of a ghat, as well as in the sense of what is collected from the son-in-law for the sake of marriage.) In this half line of the verse only one word, namely, śulka is treated. Sometimes one word is treated as the entire verse. Thus

(10) śikā jvāla śikā cūḍā śikā śākā śikā śīpha /śikā śikhaṃdīnam cūḍā śikā syād agramātrakām //

This verse is devoted entirely to the word śikā and its meanings.

In “6”, homonyms are arranged in the order of the last syllable and not of the last consonant. This lexicon is seen to be influenced by Southern vernaculars. Thus in this lexicon, words with one akṣara are listed first then with two and three and then four. Amongst them again words are arranged according to the alphabet. The words kuṭēha (kuṭēha in kannada meaning dwarf), prāya (prāya in kannada meaning age), saēa (saēa in kannada meaning small) and koēa (koēa in kannada meaning room) are used in this lexicon in the same sense in which they are used in kannada. Normally these traditional lexica are constructed in anubhūbh metre, however, “5” is constructed in more than 15 metres like vasantatilaka, caśīnē, mālinē, etc.

“9” is a specialized homonymous lexicon restricted to the names of plants, trees, and herbs that go to form the Materia Medica in the Ayurvedic system of Medicine and it records about 2860 principal and about 4860 words denoting the meaning thereof. There are almost 44 names of authors as well as works in the field of Ayurveda and about 44 works and names of authors of general lexica are found quoted in this lexicon. This lexicon provides the names of the original homelands of the medicines mentioned. Thus, words like criāṛṛē, gambhāre, kaōphāl, hirā, etc., which indicate names of the plants, are stated to be originated in the region of Kashmir. Kalinga is mentioned as the homeland of lāigāla, kuōja, gaura, etc.

“10” is a very peculiar type of lexicon, perhaps unique of its kind. It mentions words of only one and two syllables. Thus-

(11) kuū pāthvē kuū kucaū kulam kuū kātyā bhūrapīsmātā/

This line records words with one vowel with k consonant at the beginning with their meanings. They have not been recorded by the modern Lexica.

In “11”, following features can be highlighted: a) the homonyms are arranged into groups according to the initial letters and the number of syllables in each letter. b) in the section of Synonymous words a desī word is also recorded. c) Vernacular phrases are tagged before each set of synonyms.

Compound words are not used as far as homonyms are concerned. Cases for anusvāra are not dealt with. ‘k’ appears after ‘h’. We find-

(12) jambhēro jambhālo jambho jambhēro deçayasamgrahe/

(13) pēlakhitindolite cāpi decoye hinīcitam iñyate //

In a nutshell we can summarize the features of the traditional lexica as follows. Features of this phase are:

- The alphabets were ordered systematically according to phonetic principles, to this corresponded the macrostructure of the lexicon. However, alphabetical order, as far as the beginning of the word is concerned, was not followed.
- Lexicography was developed systematically. The stem (or root) was used as the lemma.
- A theory of eponymy was devised.
- A theory of compounding was developed.
- Dialect differences were noted.

F. The Concept of Homonomy and Synonomy

It is worthwhile to study a little about the concept of Homonomy and Synonomy that these lexicographers used. The Synonymous lexica include both simple as well as compound words, whereas the homonymous lexica include only simple words and rarely the compounded words.

The concept of Synonomy is not very well defined by these traditional lexicographers. As Ghatge(1973:28), points out, “The only criterion by which they would like to use appears to be the idea of parivātisahava or the possibility of replacement. This is taken by them as a purely semantic criterion and a difference of gender alone is not sufficient to destroy the quality of paryayāvata. We are thus left with their practice to see what they mean by a synonym.” Words are classified into three categories, niūha (Conventional), yaugika (derivative) and yogyāniḥa (derivative but restricted in usage by the convention). There are various kinds of relations which lead to
syonymy as listed by the traditional lexicographers. As far as the Homonymy is concerned there are several problems. The traditional lexicographers do not define the concept of anekārtha-phada. At the same time the concept of Polysemy and Homophony remains undecided for these lexicographers.

IV. MODERN LEXICA OF SANSKRIT

The first Sanskrit Dictionary with western system of alphabetical order was the Sanskrit-English Dictionary compiled by Professor Horace Hayman Wilson and published in 1813. Two Indian works, viz. the Saubdakapdruma compiled by Pandit Sir Raja Radhakanta Dev and the Vaasapya compiled by Pandit Taranatha Tarkavacapati, followed suit.

A. Wörterbuch

But it was the Sanskrit-German Dictionary, Sanskrit Wörterbuch, compiled by Otto Böhtlingk and Rudolph Roth, published from St. Petersberg during a span of 24 years-1852 to 1875 - which is considered to be the benchmark as far as the lexicography related to Sanskrit is considered. The arrangement of meanings on historical principles is the methodology followed by this lexicon, which we see is adhered to by all later lexicographers. It drew its material from 450 books.

B. Monier Williams

This dictionary was compiled by Monier-Williams. Though he acknowledged his indebtedness to the Sanskrit-Wörterbuch, he worked for his dictionary on a plan of his own. It contained several features which had not been found in the Wörterbuch. One can have an idea of those features from the subtitle of the dictionary, 'etymologically and philologically arranged, with special reference to cognate Indo-European languages.' The first edition of the dictionary was published in 1872 and the author, as soon as he became aware of the likelihood of his volume becoming out of print, set about preparations for a new improved and enlarged edition. He revised his original work in view of the criticisms, which the first edition invited from competent scholars and many Sanskrit texts that had been coming to light since the publication of the first edition. The second edition contains words and derivatives culled from more than four hundred Sanskrit texts. The author, while recording different shades of meaning of the words, has taken the meanings given by Indian lexicographers Kosa-karas and also the contexts of actual places of occurrence in the texts into consideration. The developments of meanings have been shown in chronological order. The author has also utilized the results of the research in comparative philology as available in his time. It drew its material from 515 books.

The main feature of these two lexica is that they record the lexemes in the forms of stems. They also provide ample evidence in the form of the references where a meaning of a particular word occurred. ‘kri’ occurs as a conjunct consonant at the end of ‘k’. Negative compound is treated as independent lexeme. Compounds with prepositions are mentioned under that preposition. Thus the words upakāra, upakāt, etc. are found under upa. The immediate constituent cut is the essential feature of these two lexica. These are the lexica of both nominal as verbal stems. The derived roots are mentioned under the original roots. Thus the desiderative roots like sikāra are mentioned under kā. However, not all such derived roots are treated in these lexica.

C. Encyclopaedic Dictionary of Sanskrit on Historic Principles

This is an encyclopedic dictionary drawing its lexemes from 1500 works, thrice more than that of the Monier Williams. It records lexemes from Vedas up to the end of 18th century. The basic aim and plan of this dictionary seems to be the following-

(i) To indicate the earliest and the latest (in case it has become obsolete) occurrence of the vocabulary in Sanskrit Literature.

(ii) To indicate whether a word existed throughout the history of that language or was confined to certain periods.

(iii) To show the provenance of the word, i.e. whether it is used in many branches of learning or if it is system specific.

(iv) To show all available meanings, common as well as technical, and by arrangement, to show their historical relationship if traceable.

It is an important question as to how these lexica are related to Sanskrit grammar and more specifically to the grammar of Panini? How does Panini treat them for different grammatical processes? But this will not be dealt with in detail in the present paper.

V. GRAMMATICAL LEXICA

It is important to note that there are strictly system specific lexica in Sanskrit, significant among them are the lexica built for the purpose of grammar. The most important among them is the Lexicon of verbal roots or dhatupatha as it is traditionally known. This is believed to have been composed by Panini himself and awaits attention for its scientific arrangement from the point of view of grammatical functioning. The lexemes are arranged in ten groups according to the morphological operation they get in some specific environment. These groups are called as gātas. Lexemes in each gāta are further classified according to the application of certain types of suffixation. Based on this, the lexemes are
called "...padin"s. (ātmāpadin, parasmaipadin and ubhayapadin). Each lexeme is enjoined with a marker to indicate certain types of suffixations, certain augments, certain modifications and other grammatical operations. The rules in the grammar of Panini are aware of these classifications and this arrangement. They often refer to the sub-classification amongst these gānas. We can say that this lexicon is very much system specific and leads to generation of specific forms. The rules related to these lexica generate adjectives and can be used to denote anything which performs a particular action. This is supported partly by the theory recorded by Yaska that all nouns are derived from verbal roots.

As far as the lexica mentioned above deal with the nouns, Panini's grammar reclassifies all the nouns on formal basis. He classifies them on the basis of the final element of the word. Thus the first classification is words ending in vowels and those ending in consonants. We do not find this kind of classification in any of the lexica.

Panini's grammar talks of Karaka relations and the generative aspect of this grammar is closely related to the Karaka theory. With the help of this theory words can be generated from a given lexeme which will indicate the relation to a particular action.

Below given is a table in which a sample number of forms are shown to be generated with the help of grammatical rules of Panini from the lexicon of verbal roots, which can be used for coming new terms for scientific and technological purposes.

<table>
<thead>
<tr>
<th>Root</th>
<th>Suffix</th>
<th>Karaka</th>
<th>Term</th>
<th>Meaning</th>
<th>Rule</th>
<th>Comment</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>krit</td>
<td>kām</td>
<td>ādayākār</td>
<td>One Word done</td>
<td>3.1.122</td>
<td>Adj</td>
</tr>
<tr>
<td>Ts</td>
<td>kāra</td>
<td>kāra</td>
<td>ādayākāra</td>
<td>One Which is done</td>
<td>3.2.108</td>
<td>Adj</td>
</tr>
<tr>
<td>Ad</td>
<td>krit</td>
<td>kām</td>
<td>ādayākār</td>
<td>The instrument of doing</td>
<td>3.3.117</td>
<td>Adj</td>
</tr>
<tr>
<td>Ts</td>
<td>kāra</td>
<td>kaliyādhyā/ katriyā/ kāmādhyā</td>
<td>What Should be done</td>
<td>3.1.83</td>
<td>Adj</td>
<td></td>
</tr>
</tbody>
</table>

I have not taken here the addition of the preverbs which are known as upasargas. They will be helpful in generating further meaning variations, and in fact, some new meanings altogether. In this manner any verbal root taken from the verbal lexicon can generate with the help of a grammatical rule and a suffix a number of forms.

VI. APPLICATION OF PANINI'S GRAMMAR FOR CREATING NEW LEXICA:

It is this feature of the grammar of Sanskrit as well as that of compounding that can be effectively used for creating new technical terminology for Indian languages. Here take the 30 + lexica published by the Government of Maharashtra, as a case study for illustrating this point in brief. Ever since 1967, the Directorate of Languages, Government of Maharashtra started the activity of preparing the lexica of technical terminology for all fields of Science and Technology as well as human sciences. The underlying principles for this activity are elaborately stated in the introduction to the lexicon for technical terminologies for Linguistics. It is a 19 points decree for such an activity. These points include Exhaustive reference, Relevant reference, Amenability to testing referential fit, amenability to calibration, etc.

These technical terms are two types- "technical terms collected into a technical terminology" and "technical terms collected into a technical nomenclature". For creating technical terminology of both these types verbal morphology as well as compounding features of Sanskrit are seen very effectively used in these lexica. Thus for example in the Lexicon on Linguistics, for 'acronym' the term coined is 'ādayākārā cañḍāla'. This is of the second type mentioned above. It is formulated by the process of compounding and taddhita formation (derivation process of an adjective from a noun by adding a suffix). We can show the process as follows-

\[
\text{Adyā + aikāra = adyākārānā} \\
\text{adyākārānā + in = adyākārārīn}
\]

This becomes the adjective of cañḍāla, which together with this adjective can be considered to be the coined technical term for acronym. Thus in order to coin the terminology for magnet and related words all the methods mentioned above are used. The technical nomenclature of a magnet is grasped in the word formed by adding the suffix "aka" after the root "cumb".

\[
\text{Cumb} \\
\text{Cumb + aka = cumbaka}
\]

By the process of compounding other words are added to the above word and as many words are coined.

\[
\text{Vidut + cumbaka = vidyutcumbaka (electromagnet)} \\
\text{Nala + akāti + cumbaka = nalakāti cumbaka (horseshoe magnet)}
\]

\[
\text{Cumbaka + āya = cumbakeyā (magnetic)}
\]

It is possible to add a taddhita suffix after the compound. It is also possible to compound after the taddhita suffix is added.

Thus we have following terms by the abovementioned process-

\[
\text{Cumbakeyā pravardha (magnetic amplifier)} \\
\text{Cumbakeyā eka ak ūk (magnetic axis)} \\
\text{Cumbakeyā pariṣṭha (magnetic circuit)} \\
\text{Cumbakeyā rātri (magnetic field)} \\
\text{Cumbakeyā kātra (magnetic needle)} \\
\text{Cumbakeyā aman (magnetization)} \\
\text{Cumbakeyā amā (magnetism)} \\
\text{Cumbakeyā varati (magnetometer)} \\
\text{Cumbakeyā bala (magnetizing force)}
\]

164
In the same way, in almost all the lexica mentioned above, we find this principle used. In future as well we can make use of this feature of Sanskrit grammar in order to coin new terminologies for upcoming branches in the field of Science and Technology. We can also borrow certain grammatical elements from other languages and some elements from Sanskrit for this purpose. Thus in Linguistics, the terms “rūpīma” is coined for Morpheme and “svanīma” for phoneme. These are created out of the Sanskrit roots “rūp” and “svan” respectively and the English morpheme “eme”. In this way, keeping in mind the 19 points mentioned above it is possible to borrow linguistic elements from different Indian languages and terms can be coined.

APPENDIX

Some more Lexica (traditional as well as modern)
1. Anekarthadhyanimajjare of Mahakshapanaka (H = Homonymous)
2. Anekarthasamagrah of Manikha (H)
3. Anekarthasamagrah of Hemacandra (H)
4. Abhidhānānāmakalā of Halayudha
5. Anekarthasamuccaya of Svasata (H)
6. Trīkāṭaṭṭia of Purushottamadeva
7. Nārāthasamagrah of Ayaja (H)
8. Medineof Medikara
9. Viṣṇavakrama of Mahesvara
10. Viṣṇuvacana of Sridhara
11. Dhaturatnakara
12. Dictionary of Panini by S.M.Katre
13. Nyayakoça of Jhalakikar
14. Mēmānsakoṣa of Kevalananda Sarasvati
15. Dictionary of Sanskrit Grammar by K.V.Abhyankar
16. Sanskrit Cutilan Dictionary

BIBLIOGRAPHY

Primary Sources:

Secondary Sources:

Web links:

This paper was presented at LRL-2007: National Seminar on Creation of Lexical Resources for Indian Language Computing and Processing at C-DAC Mumbai (26th to 28th March 2007), jointly organized by the Commission for Scientific and Technical Terminology (CSTT), New Delhi, MBIR, Govt. of India and the Centre for Development of Advanced Computing (C-DAC), Mumbai, Department of Information Technology, MC&IT, Govt. of India.