Ranganathan’s Colon Classification:
Kannada-English Version ‘dwibindu vargiikaraNa’

Abstract

“dwibindu vargiikaraNa” (ದ್ವಿಬಿಂದು ವರ್ಗೀಕರಣ) is the Kannada rendering of the revised Colon Classification, 7th Edition, intended essentially for the classification of macro documents. This paper discusses the planning, preparation, and features of Colon Classification (CC) in Kannada, one of the major Indian languages as well as the Official Language of Karnataka, and uploading the CC on the web. Linguistic issues related to the Kannada rendering are discussed with possible solutions. It creates facilities in the field of Indexing Language (IL) to prepare products such as, Subject Heading List, Information Retrieval Thesaurus, and creation of subject glossaries or updating the available subject dictionaries in Kannada.

Need for ILs in Indian Languages

India is linguistically rich with 1,652 mother tongues of which 22 are Scheduled Languages (and 100 are non-scheduled languages) included in the Constitution of India. However, there is a paucity of indexing languages (IL) in Indian languages (Sharada, 2010). Most of the existing IL are in English and a few other languages like Chinese, French, German and Italian. Realizing the importance of IL in Indian languages, M.A. Gopinath had written in the Introduction of the 7th edition of the Colon Classification (CC) that he hoped to develop the scheme in Indian languages as well, in collaboration with other organizations (Gopinath, 1989). The history of documents published in regional languages of India goes back to the early 19th century. For instance, the first book in Kannada was published in 1817 (Wikipedia). There are innumerable concepts native to a particular region, language and culture which are not found in the English language schedules. Tools to properly organize documents on such concepts in a user-friendly sequence are lacking and therefore, there is a need to prepare IL in Indian Languages.

CC in Kannada and other Indian Languages

The first rendering of the CC 7th English Edition into an Indian language is “dwibindu vargiikaraNa” prepared and published through the collaborative efforts of Sarada Ranganathan Endowment for Library Science (SRELS), Bangalore and the Central Institute of Indian Languages, Mysore (CIIL) in 2010. Udaya Narayana Singh, former Director of CIIL, in his foreword to this volume, states that:

"On the event of Kannada getting the Classical status, with its richness of literature in all walks of life, what was lacking was the tool to properly organize the universe of knowledge. Hence, as a test case, Kannada was accorded priority."

This has opened up opportunities to develop several information retrieval tools in Kannada. This is the first bilingual classification schedule adequately updated and flexible enough to accommodate any number of new concepts.

Linguistic Features Across Language Families

A major linguistic discovery of the 20th Century relating to Indian languages is the identification of common linguistic features across different language families. Indian languages belong to different language families but do share common linguistic features, hence India is referred to as a „linguistic area” (Emeneau, 1956). This sharing of linguistic features across the language families was facilitated by their coexistence for centuries together, and by the continuing interaction between the people speaking these languages.
The language families are: Indo-Aryan (Hindi, Bangla, Marathi, etc.), Dravidian (Kannada, Malayalam, Tamil, Telugu, etc.), Tibeto-Burman (Khasi, etc.), Astro-Asiatic (Santhali, Mundari, etc.) and Sino-Tibetan (Bhutanese, etc.). A 19th Century missionary to India, Rev. William Campbell wrote (1839):

“Whatever may be the difference in the languages, they all belong to the same great family; similar laws regulate the idiom, construction, style, and various kinds of composition, which prevail in the dialects of the north and the south; when you describe one art of India, you have, in many respects, described the whole: the manners, the customs, and the habits of the people, with trifling variations, correspond from Cape Comorin to the Himalayas: and their superstition, in all its great lineaments, is exactly the same. Whether, therefore, their present literature was originally written in Sanskrit, or in some other languages, the Vedas, the Shastras, the Puraanas, and all their classical writings are to be found in all the principal tongues of India, and are as well understood in the one as in the other”

Languages within a group also share a number of common elements. Due to shared linguistic features like vocabulary, and grammatical elements, and common patterns followed in the creation of technical terminologies by the Centre for Scientific and Technical Terminology across the languages in the post-independence India, rendering the same into other Indian languages (with one or two exceptions) should not be a difficult task.

Some of the shared linguistic features across language families are as follows:

1. Presence of a series of retroflex consonants that contrast with dentals sounds [t/ D].
2. Two to three degrees of ‘you’. [[tu „you” (sg), tum „you” (pl) and aap (hon) in Hindi and niinu „you” (sg), niivu „you” (pl) and taavu (hon) in Kannada]]
3. Widespread lexical borrowing. (samaya, kaala, aadeesha)
4. Presence of echo word constructions and onomatopoeic forms. (gaD baD, daDa)
5. Reduplication process of different grammatical categories such as nouns, verbs, adjectives, adverbs, etc. (baDe baDe, doDDa doDDa)
6. Compound verb forms.
7. Conjunctive particle. (uu „also”)
8. Sentence structure - flexibility of word sequence though finite verb usually comes in the last position. (avaru manege baruttaare, vah ghar aayeegi)

Relevance of CC

S. R. Ranganathan’s Colon Classification (Ranganathan, 1963) based on his theory of freely faceted classification has a modular structure and is sufficiently flexible to enable the addition of new concepts without the structure being seriously disturbed (Neelameghan, A., 2000).

Planning and Preparation

Data was collected from cc 7th Edition and broad translation was done. Each descriptor in the schedule in Kannada was rendered after checking subject dictionaries/glossaries available in many forms such as print, manuscript, and digital. Subject experts with proficiency both in a particular subject as well as the Kannada language, were involved to assist in this venture. Chapter headings, representation of Basic subjects within a main subject, notation, descriptors, expression of rules (example: Divide by CD), etc, were planned before preparing the schedules so as to ensure consistency and uniformity in the entire schedule.

Considering that many developments had occurred since the 7th edition of cc was published, the schedule of Basic subjects in the English version was revised. New concepts were introduced without disturbing the existing ones wherever necessary. For example: BX Astronomy: (Ranganathan, 2010, p.167):
The book has six chapters. Chapters 1 to 3 provide an introduction to CC, its postulates and principles. Chapter 4 lists traditional main subjects, schedule of basic subjects, and schedules for languages and common isolates such as Time, Space, Energy, Matter Property and Personality facets. Chapter 5 provides schedules of special isolates for various Basic classes. Chapter 6 is the Index and the list of books referred.

**Common Isolates**

In the schedule of languages, due importance is given to Indian languages, with special reference to the languages mentioned in the 8th Schedule of the Constitution of India. Similarly in Space schedule, the districts in the state of Karnataka are fully listed. Common schedules for Matter Property and Energy isolates have been expanded.

**Special Isolates**

The „E Chemistry“ schedule of Inorganic elements, 16 groups are listed as against 8 in the English edition. In „I Botany“ and „K Zoology“ schedules, based on International Code of Botanical Nomenclature (ICBN) and International Code of Zoological Nomenclature (ICZN) respectively scientific names in Botany and Zoology are Latinized. Its intent is that, each taxonomic group of plants has only one name accepted world wide. Many plants of Indian origin are included. Arts and crafts related to Karnataka are also included. Newly added subjects are:

- 5.9B ಕಂಪ್ಯೂಟರ್‌ದ ವಿಜ್ಞಾನ - Computer Science
- 5.AB ಪರಿಸರ ವಿಜ್ಞಾನ - Environmental Sciences
- 5.AC ಪರ್ಯಾಯವಿಜ್ಞಾನ - Surface Science
- 5.3X ಗುರುತಿತು ಪರ್ಯಾಯ - Book Publishing
- 5.LZ ಕ್ರೀಡೆಗಳು – Sports

**Index**

Many concepts at micro level not included in the schedules may be found in the index. The present work is for classifying macro-documents. Instead of elaborating some of the concepts in the main schedule, they have been included in the alphabetical index.

**Special Features**

CC is facet based and recognizes five Fundamental Categories (FC) as mentioned in Table 1.

<table>
<thead>
<tr>
<th>Name of FC</th>
<th>Indicator Digit</th>
<th>Abbreviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Personality</td>
<td>,</td>
<td>P</td>
</tr>
<tr>
<td>Matter Material</td>
<td>;</td>
<td>MM</td>
</tr>
<tr>
<td>Matter Property</td>
<td>;</td>
<td>MP</td>
</tr>
<tr>
<td>Energy</td>
<td>:</td>
<td>E</td>
</tr>
<tr>
<td>Space</td>
<td>.</td>
<td>S</td>
</tr>
<tr>
<td>Time</td>
<td>,</td>
<td>T</td>
</tr>
</tbody>
</table>

The Facet formula BS, [P]; [MP];[E];[S];[T] is applicable to all the schedules. The abbreviations for the five FCs mentioned in the above Table 1 and the five devices
mentioned below are retained as they are, since these devices are more familiarly known in the abbreviated form in English.

- **CD** Chronological Device
- **GD** Geographical Device
- **SD** Subject Device
- **AD** Alphabetic Device
- **ED** Enumeration Device

Throughout the schedule these devices are used as mentioned below:

(AD ಕಲ್ಲು ವಿಭಾಗದಿ - Division by AD)

{ದ್ವೇಷಾತ್‍ರಾಯ - Example}

zG  ಗಢಿ ವಿದ್ಯಾರ್ಥಿ - Gandhiana

**Linguistic Issues Related to Rendering in Kannada**

a) For a single English descriptor different terms are found in different dictionaries. While taking into consideration the context, one has to depend on the entry in CC schedule based on the subject.

For example: Male – gandu, ganDasu, purusha, manushya
Though all the four terms are in use, the word most suitable to the subject has to be chosen for representation.

b) Since a few descriptors do not have exact Kannada equivalents, they have been adopted as they are in English (transliteration). Schedules for Space Isolates, Botany and Zoology are good examples. The major problem with transliteration is for the words beginning with Vowels such as E, I and O and consonants – J, C and K. In textual transliterations it may depend upon the context. As we are considering only the concepts, based on the phonetics, equivalent terminology is provided. For the pronunciation, standard dictionary and the internet site http://www.answers.com is referred.

For the term Engineering:
1. ಇಂಜಿನಿಯರಿಂಗ್ inginiyaring [University of Mysore English – Kannada Dictionary]
2. ಇಂಜಿನಿಯರಿಂಗ್ enginiyaring [Vijnaana Kosha]
Both are in use, but we have to adopt one. The latter is chosen since, in the transliteration, the letter “e” represents Kannada letter “ə” [ex: ettara əತ್ತರ] and “i” represents “ಇ” [ex: ಇಂಡಿಯ india].

In some cases, borrowed words from Indo-Aryan languages into Dravidian with localization are chosen. For example: Manager – nirvaahak, prabandhak are the Indo-Aryan equivalent. It is transformed into Kannada by adding an inflection “a” and the Kannada word is nirvaahaka or prabandhaka.

c) Derivational implications

The term “Chemistry” has two renderings ರಾಸಾಯನಿಕ saastra and ರಾಸಾಯನಶಿಲ್ಪ rasaayana Vijnaana in two different dictionaries. The latter is determined to be appropriate since, in Kannada, if “-ಅಕ್ಕ” (ika) pratyaya /inflection is added to a root word, the first syllable of the word becomes long. Several examples are there for this group.

ಸಾಮಾಜಿಕ samaaja (Society) - ಸಾಮಾಜಿಕ saamaajika (Sociological)
Web Interface
In December 2011, a website on dwibinduvargiikaraNa was developed (accessible at http://www.classicalkannada.org.) Here, the full text of preliminary pages and first three chapters can be found. For the schedules of common isolates, and special isolates sample page link is provided including the index.

Uses of DwibinduvargiikaraNa
It is helpful to library professionals who are educated in Kannada medium throughout their educational career.
1. There are several concepts native to a particular region and its culture. Folk arts of Karnataka are a good example.
2. The elaborate alphabetical index is useful in formulating subject headings in Kannada, as the basis for the development of an information retrieval thesaurus as also lexical tools similar to WordNet, technical glossary/dictionary, etc
3. The present work “dwibindu vargiikarana” can be used as a base for translating CC into other Indian languages
4. Subject experts who helped in this work have also found it useful for them to teach their subject in Kannada. The hierarchical arrangement of concepts is very useful to them.

Future Agenda
1) Based on the dwibinduvargiikarana, essential information processing and retrieval tools in Kannada, such as thesauri (IRT) and lists of subject headings are being planned. A bilingual thesaurus (Kannada- English) is in progress. As a test case the discipline “Education” has been selected. The related thesauri available online is being used for English and are rendered to Kannada with class numbers from CC.

The need for a list of subject headings in Kannada was much felt when - library retro conversion work was in progress. The cataloguing of books in Indian languages was done in both the language of the item and in English except for the subject heading field which had to be only in English due to the non-availability of such tools in Indian languages.

2) Natural Language Processing (NLP) application
The development and applications of NLP in Indian languages are at various stages of development. In a recent survey of publications between 2000 and 2010, 297 papers on development and application of NLP in information retrieval in Indian languages were found of which only 14 papers were on Kannada which stands fifth in the rank list (Sharada 2012).

Earlier studies in English in this area tried to find the suitable grammar worked out for natural language. It is useful to explore the feasibility of keeping the five fundamental categories as the delimiters to develop parsers for Indian languages. Separate files need to be prepared for Basic Subject (BS) and the five FCS with an algorithm. After developing this, the schedule can be used online with the flexibility to instantly incorporate new developments in the respective disciplines.
References


Emeneau, M.B. 1956. India as a linguistic area. Language, 32:3-16.


