

## 7.0 Standardisation – Unicode for Indic Scripts

Unicode Standards are widely being used by the Industry for the development of Multilingual Softwares. The basic Input for evolving the Unicode Standard for Indian scripts were taken from “Indian Script Code for Information Interchange-1988 (ISCII-1988)” document, whereas the present national standard is ISCII:1991 (ISCII- IS13194:1991). Hence some discrepancies observed in the Unicode for Indic Scripts. Department of Information Technology, Ministry of Communications & IT, became the voting member of the Unicode Consortium to adequately represent Indian scripts in the Unicode Satnadards. Some modifications were necessary to incorporate in the Unicode Standard for correct and complete representation of Indian Scripts. Department of Information Technology organized number of meetings with concerned State Government IT Departments, Linguist and Indian IT experts. DIT finalized the proposed changes in the Unicode Standard in consultation with respective State Government, Indian IT Industry & Linguists. The proposed changes were published in TDIL Newsletter Vishwabharat@tdil issue 4 (Devanagari based languages Sanskrit, Hindi, Marathi, Nepali, Konkani & Sindhi), issue 5 (Gujarati, Malayalam, Telugu, Gurmukhi, Oriya), issue 6 (Bangla, Manipuri & Assamese), issue 7 (Tamil, Kannada, Urdu, Sindhi & Kashmiri). The proposal was submitted to Unicode Consortium. Unicode Technical Committee (UTC) has accepted some of the proposed changes for inclusion in the Unicode Standards. These changes have been incorporated in Unicode Standard 4.0. The summary of the proposed modification in UNICODE standard for Indian scripts and their status is presented here:

**Devanagari:** In the Unicode Standard 4.0 Devanagari is represented very adequately. Department of IT proposed to include Two Marathi Characters sha, la, and Half Ra (Marathi Eyelash Ra), Devanagari Conjuncts Ksa, Jnya, Shra, Four Sindhi implosives and Three Vedic Sanskrit characters in the Devanagari code block. The Marathi characters are the allograph of Hindi characters hence these have not been assigned

separate code points. A detailed description is provided about these allograph in the Unicode 4.0. The additions of conjuncts is not encouraged as these can be realized using the rendering mechanism. The addition of Vedic Sanskrit characters is being taken separately and DIT has demanded to include Vedic code Block in the Unicode. The annotations suggested for Halant have been added at code point U+094D.

### Bengali:

In the Bengali code block two characters “Bengali Sign Avagraha” and “Bengali Khanda Ta” have been added in the Unicode Standard. Some annotations also have been added to the respective code points in the Unicode for Bengali. DIT also proposed to encode “Assamese letters RA and WA”. These characters are encoded in the Bengali code block and to address searching and sorting issues collation algorithms may be used with applications in Assamese and Manipuri. The proposed addition of Bengali Ya Phala is under consideration of the UTC.

### Gurmukhi:

Gurmukhi Visarga has been added in the Gurmukhi code block of the Unicode Standard. Gurmukhi Sign Khanda and Ek Onkar have added at code points U+0A74 and U+262C. The glyph of Gurmukhi letter LLA has been corrected.

### Gujarati:

Gujarti Currency Sign, Gujarati Letter Vocalic L, Gujarati Letter Vocalic LL, Gujarati Vowel Sign L and Gujarati Vowel Sign LL have been added in the Gujarati code block. The glyph for the Gujarati sign Avagraha has been corrected. The addition of Gujarati Fractional Numerals “PA”, “ADDHO” and “PONO” is under consideration of the UTC.

### Kannada:

Kannada Sign Nukta and Kannada Sign Avagraha have been added in the Unicode for Kannada. Also annotation has been added for the Kannada Sign Halant. The diacritic marks desired in the Kannada block may be borrowed from the Devanagari code block.

### Malayalam:

The annotation has been added for the Malayalam sign Chandrakkala. DIT also proposed to add five “Malayalam Chillakshram” in the Unicode for Malayalam.

### Oriya:

The “Oriya letter WA” have been added in the Unicode for Oriya. The glyph of the Oriya digit zero has been corrected. The annotation s proposed for some code points have also been accepted.

### Tamil:

Annotations proposed for Tamil characters have been added in the Unicode for Tamil. However, Government of Tamil Nadu proposed to have All Character Encoding scheme for Unicode for Tamil.

### Telugu:

Annotation has been added for the Telugu Sign Halant. DIT also proposed to include Telugu Sign Nukta, Telugu Sign Avagraha and Diacritic Marks.

### Unicode for Vedic Sanskrit:

DIT proposed to include Vedic Sanskrit in the Unicode Standard. A proposal for the same was submitted to Unicode Technical Consortium. UTC is considering the proposal and has asked to submit details about the proposed characters.

### Unicode for LEPCHA

#### Introduction

Lepcha Language or Raong Aaring is the language used by 3,00,000(three lakhs) Lepchas of Sikkim, Darjeeling, Kalimpong, Kurseong, Mirik, Illam (Nepal) and some parts of South West Bhutan. It has existed since time immemorial. In fact Lepcha language is the ancient language of Mayal Lyang, the land of Eternal abode. The language is a monosyllabic in nature (though not altogether an isolating one). Although non-Lepcha scholars/researches have opined its original is Mongolian Semitee or Indo Burmese etc, but Lepchas believe it is independent, ancient and distinct from any other language of the world. Lepcha

is simple and soft-spoken language.

#### Description:

The most important features of this language is that it is systematic in naming of animals, trees, rivers, etc. Almost all names of main rivers commence with the Lepcha letter “Ra” like Rungnyoo, Rungeet (Tista Rangeet) Rungyong, Ruyaot, Ruthaong, Romam, Rushi, and Ruyaon Rungtsong etc. The names of trees starts with the root letter “Sa” such as Sunglee Koong, Salieem Koong, Sunglaok Koong, Samaal Koong, Sambraang Koong and so many others. Names of wild animals commence with letter “Sa” such as Sazaong, Sazaor, Sothaon, Somee, Solaa, and Soreeng etc.

Another features of Lepcha language is “Tungbor Aareng” at figurative or indirect language. This part of speech is used by Lepchas in special cases like hunting, fishing, Collecting wild honeybees, edible foods in jungle and in a long journey.

#### History of Lepcha Language:

Lepchas strongly believe that original source of language are Munand Baong theeng (Lepcha priestess and priest). The language of Mun and Baongtheeng in Lepcha world is not in a written form. It handed down from generation to generation since time of immemorial.

Lepcha was an official language is Sikkim and Darjeeling hills till end of nineteenth century. In mid twentieth century much literary works came out from Kalimpong, which helped to review Lepcha Language in Sikkim. After Sikkim become a part of India, the language got state official language recognition in 1977 with other Languages. In the subsequent year it got its place in schools as one of the vernacular subjects and in 1982 it was introduced up to class XII. At present, this language is taught up to under graduate level in the colleges of Sikkim as Modern Indian Language (MIL).

The human Resource Development Department of Govt. of Sikkim has a Text Book Unit especially for Lepcha Language. This Text Book Unit is responsible for preparation of School Syllabus Text Books in Lepcha from pre primary to Class XII.