

Nepali Font Standards

White Paper

Version 2

1998

Nepali Font Standardisation Committee

GPO BOX 956

Kathmandu, Nepal

Tel. : 977 1 249668/249151

Fax. : 977 1 249057

E-mail : font@unlimit.mos.com.np

NEPALI FONT STANDARDS

White paper

Introduction

This document describes a proposed standard for the encoding of all the languages of Nepal within the computer. This standard has arisen out of a need felt by a number of Nepali computing professionals and users. Similar initiatives have been taken in the past, but these came to nothing - the time was not yet right for them; but now it is.

To ensure that the right decisions have been made for Nepal, a number of experts from within Nepal have been drawn upon, serving within committees as follows:

Standardisation committee, to make sure that the standard fits the policies and practices of Nepal: representatives from Computer Association of Nepal, Kathmandu University, Ministry of Information, Ministry of Science and Technology, National Computer Centre, Nepal Bureau of Standards, Nepal Press Institute, Nepal Telecommunications Corporation, Press Council, RONAST, Royal Nepal Academy, Tribhuvan University.

Language committee, to make sure that the proposal was correct for the languages of Nepal: Bairaga Kainla, Kamal Mani Dixit, Krishna Chandra Singh Pradhan, Madav Pokhrel, Yogendra Yadava.

Technical committee, to make sure that the proposal was technically feasible: Muni Shakya

The day to day development of the standard has been done by a working committee coordinated by Allen Tuladhar with secretary Gaurab Raj Upadhaya and members Bhanu Pathak, Jeff Rollins, Kanak Mani Dixit, Patrick Hall, Peter Malling and Sunil Shrestha.

This white paper has been written so that implementers and computer users can prepare for the introduction of this standard, developing new fonts and software in conformance to it. At the same time this white paper is being sent to key organisations external to Nepal, notably key computer manufacturers and standards bodies.

This white paper is divided into three major parts -

Part I which fills in the background to the standard, surveying the languages of Nepal and their needs and the needs for software working in the languages of Nepal.

Part II which defines the standard itself - the internal codes, the rendering requirements and standard glyphs for display, and the input requirements and standard keyboards.

Part III which gives guidance to implementers supplementary to that available from the Unicode Standard Version 2.0 and True Type Open from Microsoft and Adobe.

Note: The 4 fonts needed to view this document in a word processor are Gorkhali Nepali, Gorkhali Sanskrit, Annapurna, and Sabdatara. The PDF version requires Acrobat Reader 2.1 or higher.

Part I Background to the Standard

1. The languages and writing systems of Nepal

Nepal has a tradition of spoken languages and their writing that goes back thousands of years. These arose as part of the general development of civilisations and cultures in South Asia from the first millennium BC, then through many changes brought about by external invasion and influence, to the present day. It is easy to lose sight of Nepal's distinctiveness when seen in relation to its giant neighbours India and China, the most populous countries of the world containing between them more than half the population of the earth. But Nepal does have a distinct identity.

Nepal itself has only 22 million people in 1998, but is very diverse with 70 languages or dialects (Toba 1992, Malla 1989b), many of them unwritten until recently, but some with writing that goes back more than a thousand years. Table 1 shows the more significant of these languages, dividing them into major language groups, showing the number of speakers from the 1981 census, and the percentage of the population. Many of the languages have only a few thousand speakers or less, with many of these small population languages being of the Tibeto-Burmese group. There are also a few speakers of the Austro-Asiatic language Satar, and of the Dravidian language Dhangar.

Table 1. The number of speakers of the major languages of Nepal (source National Research Associates, projection for 1996 from 1981 and 1991 censuses)

Total Population = 20,055,632

Tibet-Burmese Group of Languages

Language	Number	%age
Gurung	252,381	1.26%
Limbu	28,224	0.14%
Magar	476,445	2.38%
Newari	764,067	3.81%
Rai-Kirat	486,464	2.43%
Sherpa & Bhote	134,894	0.67%
Tamang	1,001,533	4.99%
Total	3,144,008	15.68%

Indo-Aryan Group of Languages

Language	Number	%age
Awadhi	414,849	2.07%
Bhojpuri	1,527,805	7.62%
Dunuwar	26,267	0.13%
Maithili	2,427,161	12.10%
Nepali	10,301,376	51.36%
Rajbangsi	94,741	0.47%
Tharu	1,100,010	5.48%
Total	15,892,209	79.24%

Other languages

Total	1,019,415	5.08%
--------------	------------------	--------------

All writing systems of South Asia have been derived from the Brahmi system created around 2,300 years ago. Brahmi and its derivatives are alphabetic writing systems; the other alphabetic systems are the Roman system used for Western European languages, the Cyrillic system used for Russian and other languages of Eastern Europe, and the Perso-Arabic system for Arabic and other languages in West Asia. The other kind of writing system is the ideographic system used for Chinese and Japanese (though Japanese also can be written in their Kana alphabet or syllabary).

From the Brahmi base the languages of South Asia and neighbouring areas have evolved their own different writing systems which today look very different from each other. So Tibetan, Nepali, Newari, Hindi, Tamil, and even Thai, are written in Brahmi derived scripts. All these scripts have preserved a strong relationship between the way the language is written and the way it is spoken, so that the scripts are largely phonetic. Because different languages use different sounds, the scripts do have distinct and important differences.

Some of these scripts are deceptively similar to each other, and the Devanagari system for Hindi is similar to that of Newari and Nepali. But they are different not just in style and superficial appearance, but in the very essence of the writing, the letters of which it is composed. This is why it is difficult, some say impossible, to write a language in the writing system of another language - so for example, Newari cannot be written adequately using the unmodified writing system for Nepali.

Over the past 15 years it has become necessary to put these writing systems into the computer. Initial attempts to do so have not been satisfactory, and hence this current standardisation proposal. In making this proposal it has been important that the nature of the various writing systems in use in Nepal are understood and agreed upon.

2. Use of computers in Nepal's languages.

There are four broad areas of potential use for Nepal's languages in computers.

Firstly, publishing. Much printed material that we see on paper is nowadays produced using the computer. This should lead to better quality and cheaper publishing, and the use of computers has certainly helped produce more printed material in languages like Newari, Tamang and Limbu. Our national newspapers are also produced using the computer - and the readers of these papers will be very aware of the lack of quality in handling Nepal's languages. Sometimes the problems are subtle, with diacritic dots positioned incorrectly beneath a main character, and sometimes the problems are quite gross, such as the use of the wrong form of conjunct - all these problems are caused by inadequacies in the way Nepali and other languages are handled in the computer. Today you cannot publish without computers, and the quality of the result is only as good as the quality of the representation of the language in the computer.

Secondly, much information is stored and used within computers, and where this information is about Nepal and is intended for use by the citizens of Nepal, clearly it would be better stored in the appropriate language of Nepal. An example here are the bills from a major utility company of Nepal, where the subscriber's name has to be transliterated into English and the whole bill is in English apart from a small amount of preprinted Nepali. In the current state of the technology, doing this in Nepali would be possible, but would be risky due to the proprietary nature of current fonts and encodings.

Thirdly, information stored in computers may need to be transferred to other computers. This happened informally when people work together to produce this white paper; and we could do this successfully because the language we are using, English, is stored in a common standardised representation for the letters, ASCII or ISO 646, and the word-processing formatting is stored in the industry standard, Rich Text Format (RTF). We need similar widely agreed standards for the languages of Nepal if we are going to be able to transfer information around the Kingdom electronically, as we might want to do when sharing information on the Internet or running a national organisation with branches throughout the country.

Fourthly, the very computers themselves interact with their human user in a language, usually English. All those menus are in English, and the manuals you need to turn to for help are in English. Wouldn't they be better in Nepali or Gurung or Tamang or Rajbangsi or other language of Nepal, in the language that the people using the computer use when talking to each other about the computer? All computer systems in Nepal, in banks, in supermarkets, and in hotels, operate in English. Of course the Nepalese peoples are very good at languages and at English, but that should not mean that they should be forced to use English. To make these systems work in Nepali requires support from within the operating system, working to some agreed standard.

3. Current state of Nepalese languages in computers.

Currently there are a large variety of fonts for Nepali available for PCs in Nepal, and some of the very first representations of Devanagari in the computer happened in Nepal. If the hundreds and thousands of fonts available for Devanagari in India are also drawn upon, it might be thought that Nepal has everything it needs. There have also been special fonts produced for Newari, and Kirati and Limbu.

However, as seen in the previous section, the quality is not there. These fonts don't even work for Nepali satisfactorily, let alone the other languages of Nepal. And if we produce a document in Nepali on one computer, and then put that file on a floppy disk and take it to another computer, we may just find that we see garbage when we look at the document on the other computer, because the internal coding of the font on the second computer is different from that on the first computer. Table 2 shows the character sets and encodings for two popular PC fonts.

Table 2: the characters sets and their encodings for Sabdatara and Annapurna.

Sabdatara

	0	1	2	3	4	5	6	7	8	9
30				ॐ	ॐ	ॐ	ॐ	ॐ	ॐ	ॐ
40	ॐ	ॐ	ॐ	ॐ	ॐ	ॐ	ॐ	ॐ	ॐ	ॐ
50	ॐ	ॐ	ॐ	ॐ	ॐ	ॐ	ॐ	ॐ	ॐ	ॐ
60	ॐ	ॐ	ॐ	ॐ	ॐ	ॐ	ॐ	ॐ	ॐ	ॐ
70	ॐ	ॐ	ॐ	ॐ	ॐ	ॐ	ॐ	ॐ	ॐ	ॐ
80	ॐ	ॐ	ॐ	ॐ	ॐ	ॐ	ॐ	ॐ	ॐ	ॐ
90	ॐ	ॐ	ॐ	ॐ	ॐ	ॐ	ॐ	ॐ	ॐ	ॐ
100	ॐ	ॐ	ॐ	ॐ	ॐ	ॐ	ॐ	ॐ	ॐ	ॐ
110	ॐ	ॐ	ॐ	ॐ	ॐ	ॐ	ॐ	ॐ	ॐ	ॐ
120	ॐ	ॐ	ॐ	ॐ	ॐ	ॐ	ॐ	ॐ	ॐ	ॐ
130	ॐ	ॐ	ॐ	ॐ	ॐ	ॐ	ॐ	ॐ	ॐ	ॐ
140	ॐ	ॐ	ॐ	ॐ	ॐ	ॐ	ॐ	ॐ	ॐ	ॐ
150										
160										
170										
180										
190										
200										
210										
220										
230										
240			ॐ	ॐ						
250										

Annapurna

	0	1	2	3	4	5	6	7	8	9
30				!	"	#	\$	%	&	'
40	()	*	+	,	-	.	/	०	१
50	२	३	४	५	६	७	८	९	:	;
60	<	=	>	?	@	अ	म	इ	ई	उ
70	ऊ	ऋ	ए	ऐ	ओ	औ	क	ख	क्क	क्त
80	ऋ	क्ष	क्ष	ख	रु	ग	ग	ज	र	घ
90	च	[\]	^	_	ड	क्क	क्क	क्क
100	क्क	उ	क्क	क्क	च	च	छ	ज	ज	ज
110	भ	क	भ	क्ष	ज	ज	ट	ट्ट	ट्ट	ठ
120	ठु	घ	हु	{		}	~		ढ	ण
130	प	□	त	...	त	त्र	थ	थ	द	द
140	द्व	द्य	ह	द्व	ध	'	'	"	"	६
150	-	-	न	न	न	प	प	स	फ	फ
160		क्र	व	ठ	प	भ	भ	म	म	ल
170	य	ट	-	च	च	र	०	०	रु	रु
180	ल	ल	॥	.	ल	व	ठ	व	॥	॥
190	॥	श	र	श	श	श	ष	ष	ष्ट	ष्ट
200	स	स	स	स	ह	ह	ह	ह	ह	ह
210	ह	ह	ह	ह	ॐ	ॐ	ॐ	ॐ	ॐ	ॐ
220										
230		ष्ट	ि	ि	ि	ि	ि	ि	ि	ि
240		ह	ह	ह	ह	ह	ह	ह	ह	ह
250										

The thing to notice in these code tables is that the same character appears in different positions, and that they encode different character sets. What we want is that all such tables include exactly the same set of characters and that these characters always appear at the same position in the code tables. Also the styles of the characters are different, but that is legitimate and indeed that is precisely what we do want, a range of stylistically different fonts so that we can choose the style that most fits our purpose.

India has also been investing much effort in the representation of its own languages in the computer, and with Hindi as the national language and each state having their own state official language, there are 17 different languages all of which are mandated for use on official business in some parts of India. Many individuals and organisations have created representations of Indian languages in the computer following the same ad hoc approach as in Nepal, with character sets and their encoding being peculiar to the supplier, and different from everybody else's.

India has also investigated its own languages and created a standard IS 13194:1991 called ISCII (Indian Script Code for Information Interchange) for these 17 official languages, but they have not looked at the needs of the Tibeto-Burmese languages that thrive in the hills and mountains of Nepal. ISCII accommodates the writing systems of all 17 languages of India within a single table, exploiting their common origin in Brahmi. An early version of ISCII was adopted by the Unicode Consortium in 1988. ISCII is required by national government but ignored by many people in practice who do things differently. What happens in India is as likely to be determined by events outside India as within, by the products of Microsoft and by the Unicode consortium. But India in its ISCII standard has produced lots of good ideas that Nepal can build upon.

The idea underlying ISCII is that the writing systems are phonetic, and the representation within the computer and the way you input it through the keyboard should be guided by this. So what should be focused upon are the pure consonants and vowels, leaving the computer to work out details of how things are written:

- the creation of conjuncts using half character glyphs or combined characters in vertically stacked glyphs,
- the position of the matras and other diacritics.

This means, for example, that the short “i” ($\overset{\frown}{i}$) is typed after the consonant, and stored after the consonant, even though for printing it comes before the consonant: the movement of the “ $\overset{\frown}{i}$ ” glyph to before the consonant is determined by the intelligence of the print rendering system. In ISCII, to represent a conjunct the component consonants are typed and stored with explicit halants to remove the implicit short ‘a’ vowels, thus “ $\overset{\frown}{i}$ ण” is rendered as “ $\overset{\frown}{i}$ ण”.

However, there are some problems with the particular approach taken in ISCII: they encode both the vowel character and matra unnecessarily, and have chosen to use the non-phonetic halant with a short “a” (अ) implied with all consonants. There has been powerful criticism from within India on these and other grounds. It is understood that during a recent review of ISCII an encoding was proposed without the halant but an explicit short “a” and with no matras. In this approach to render “ $\overset{\frown}{i}$ ण” the internal coding would be “ $\overset{\frown}{i}$ णअ”.

An earlier version of ISCII was adopted for the Unicode tables, and it is Unicode with its adoption by Microsoft and many other suppliers of software that make it important that we in Nepal take note of the ISCII approach, and adopt an encoding that exploits the intelligence available in the rendering systems such as TrueTypeOpen. However for Nepal we do have the freedom to adopt either the halant-matra approach of the original ISCII, or the explicit vowel approach of ISCII’s critics.

4. Special Features of Nepal’s languages

There are three aspects of the languages of Nepal that require special consideration.

Three conjuncts as new consonants

While it is true that Nepali uses the character set of Devanagari, some developments peculiar to Nepali have taken place. Three of the conjuncts of Devanagari - ksha क्ष, tra त्र, and gya ज्ञ - have become letters of the Nepali alphabet, placed at the end after ha ह. Grammarians of Hindi and Sanskrit would view this as wrong, that the conjuncts should be broken down into their constituent consonants, and treating these conjuncts as new consonants would be wrong for Hindi; but it is right for Nepali.

Chandrabindu and anuswar

Note that in Nepali the chandrabindu and anuswar are not really distinct. For vowel nazalisation both diacritics are used in the writing, but only as alternatives - Acharya (1991) describes “ \tilde{v} ” (which he terms ‘anuswar’) for this but notes that “ $\overset{\cdot}{v}$ ” may also be used ‘inconsistently’ (page 70), while Mitchell describes these as alternatives with the “ \tilde{v} ” above vowels that do not go above the line, and “ $\overset{\cdot}{v}$ ” for vowels that do, but terms these the opposite way round as ‘chandrabindu’ and ‘anusvar’ respectively (pages 5 and 7). Further the use of “ $\overset{\cdot}{v}$ ” to denote a general nasal consonant, the nasal consonant of the varg of the consonant that follows, Mitchell states that “there is a growing tendency to use the nasal consonant in preference to anusvar” (p16); Acharya calls this ‘sirbindu’ and notes inconsistencies in its use (page 70). For example anka (number) can be written “अंक” or “अङ्क” with the common practice in Nepal being the latter - “ङ” is the nasal consonant of the first varg to which “क” belongs.

All this suggests that for Nepali, there should only be one code for both chandrabindu and anuswar as vowel modifiers, and that which form this takes would be dependent upon the font and how it is rendered. However, Newari does require both of these, so they are both required in Nepal. It also suggests that the use of anuswar as a consonant nazalisation modifier may not be necessary and be introduced as part of the rendering system.

Glottal stops

Some languages of Nepal, like Limbu, use a glottal stop. Glottal stops are produced by the complete closure of the throat at the vocal chords (the glottis) and then its sudden release. Glottal stops occur in Arabic and related languages, but do not occur in Nepali or English (though they occur in the English Cockney accent in word like “bottle” where the “t” is ‘swallowed’ and replaced by a glottal stop).

5. *Why we need a standard, and why we need it now.*

We need a common and agreed way of representing the languages of Nepal within the computer. Everybody should work to the same representation and internal codes, so that we can move information in Nepal's languages between packages of software and between computers. We must do this now before too much information gets stored in the computer in unregulated and different ways.

Of course this does not mean that everybody uses the same font or writing style, just that when you switch from one font to another you can still read the text.

The representation of languages in computers is at a crucial stage now, with major new software expected from Microsoft working in concert with the Unicode consortium. The Unicode consortium aims to be able to represent all the languages and writing systems of the world, but currently does not recognise any of the languages of Nepal! It assumes that Nepali is written in Devanagari, and knows nothing about the other languages. Microsoft is aiming to implement Unicode in its Windows operating systems, and will do whatever is mandated by the Unicode Consortium.

Nepal must promulgate a standard now to influence these developments. Our standard must be registered with and be accepted by the Unicode consortium and then be supported by Microsoft. We must do this now.

References

- Acharya, Jayaraj (1991) **A descriptive Grammar of Nepali and analysed corpus**. Georgetown University Press, Washington DC, US.
- Bureau of Indian Standard (1991) **IS 13194 Indian Script Code for Information Interchange - ISCII**. Bureau of Indian Standards, Manak Bhavan, 9 Bahadur Shah Safar Mark, New Delhi 110002, India.
- Malla, Kamal P. (1989a) (editor) **NEPAL, perspective on continuity and change**, Center for Nepal and Asian Studies, Tribhuvan University, Kirtipur, Nepal
- Malla, Kamal P (1989b) *Language and Society in Nepal*, in Malla 1989a.
- Matthews, David, (1992) **A course in Nepali**. School of Oriental and African Studies, University of London, UK.
- National Research Associates (1997), **Nepal Record on Nepalese Development. Nepal District Profile**. NRA
- Toba, Sueyoshi (1992) **Language Issues in Nepal**, Samdan Books and Stationers, PO Box 2199, Kathmandu, Nepal

Part II The standard.

1. The alphabet to be encoded.

As seen in Part I, Nepali uses the character set of Devanagari, but has evolved three new consonants from the conjuncts of Devanagari - ksha क्ष, tra त्र, and gya ज्ञ - placed at the end after ha ह. The other languages of Nepal have sounds (phonetic values) not present in Nepali, and thus require extra letters like the glottal stop, as well as not requiring some of the letters of Nepali. The new letters might be borrowed from Sanskrit, or might be created especially for the language, perhaps from some existing Devanagari letter with some extra diacritic mark like a dot to differentiate it.

This has led to the following table of letters of the combined alphabets of Nepal, arranged in the order in which they appear in the telephone directory, dictionary, and similar documents.

Table 3. The alphabet for the languages of Nepal.

Notes. The last two vowels are required for Newari but not for Nepali, while the second palatal consonant is required for Hindi but not for either Nepali or Newari. There are two vowel nazalisation modifiers - chandrabindu and anuswar required for Newari. Anuswar as a nasal consonant has not been included, it being assumed that the correct nasal consonant will be used, and replaced by the anuswar during rendering if that is what the font does. Other languages have been accommodated by the inclusion of four vowel and consonant modifiers to be determined to enable them to match their own phonetic structures. The devanagari characters ष and ढ are required for Nepali but not Newari, and if needed for other languages the dot beneath them would be handled by a consonant modifier.

LETTER		COMMENT
Description	character	
Vowels		
nazalisation modifier chandrabindu	◌̣	
nazalisation modifiers anuswar	◌̇	Required for Newari, but not for Nepali
visarga	:	
avagraha	◌̣	Required for Sanskrit, Maithili
umlaut	◌̈	Required for Thulung language
vowel modifiers	VM1 VM2 VM3 VM4	For expansion of the codes to handle new languages
vowel letter short A	अ	
vowel letter long AA	आ	
vowel letter short I	इ	
vowel letter long II	ई	
vowel letter long U	उ	
vowel letter short UU	ऊ	
vowel letter vocalic R	ऋ	
vowel letter vocalic RR	ॠ	
vowel letter LR	ऌ	

vowel letter short E	ए	
vowel letter diphthong EI	ऐ	
vowel letter short O	ओ	
vowel letter diphthong AU	औ	
vowel letter anusvar	अं	
vowel letter visarga	अः	
Consonants		
varg 1, velar consonant Ka	क	
varg 1, velar consonant KHa	ख	
varg 1, velar consonant Ga	ग	
varg 1, velar consonant GHa	घ	
varg 1, velar nasal consonant NGa	ङ	
varg 2, palatal consonant Ca	च	
varg 2, palatal consonant CHa	छ	
varg 2, palatal consonant Ja	ज	
varg 2, palatal consonant JHa	झ	
varg 2, palatal nasal consonant NYa	ञ	
varg 3, retroflex consonant TTa	ट	
varg 3, retroflex consonant TTHa	ठ	
varg 3, retroflex consonant DDa	ड	
varg 3, retroflex consonant DDHa	ढ	
varg 3, retroflex nasal consonant NNa	ण	
varg 4, dental consonant Ta	त	
varg 4, dental consonant THa	थ	
varg 4, dental consonant Da	द	
varg 4, dental consonant DHa	ध	
varg 4, dental nasal consonant Na	न	
varg 5, labial consonant Pa	प	
varg 5, labial consonant PHa	फ	
varg 5, labial consonant Ba	ब	
varg 5, labial consonant BHa	भ	

varg 5, labial nasal consonant Ma	म	
non-varg consonant Ya	य	
non-varg consonant Ra	र	
non-varg consonant La	ल	
non-varg consonant Va	व	
non-varg consonant SHa	श	
non-varg consonant SSa	ष	
non-varg consonant Sa	स	
non-varg consonant Ha	ह	
Conjunct ksha	क्ष	
Conjunct tra	त्र	
Conjunct gya	ज्ञ	
glottal stop	?	Required for Limbu
nukta	़	Commonly used modifier for borrowed words
anuswar	ं	Nazal modifier, may not be necessary, but part of the rendering. – equivalent to the nazal consonant of the varg of the consonant
consonant modifiers	CM1 CM2 CM3 CM4	For expansion of the codes to handle new languages

Note that there are many letters in common between languages, but no language uses all the letters. It is this combined alphabet that will be encoded in the computer. There will be a Unicode table in which all vowels in their non-matra form follow their consonants and the implied vowel in the conventional writing systems is made explicit as the short “a”, अ. And there will be a “bridging” encoding for temporary use until support for Unicode becomes universally available. Simple translation paths should be available for conversion from existing TTF fonts to the bridging TTF encoding and also to Unicode when needed.

2. Sort orders.

Let us use the Unicode encoding with the characters as in Table 3 and the notes following it. Then if we take the ordering of the individual letters of the alphabet as the sequence given, the sort order for words is obtained directly from the lexicographical ordering of strings over this alphabet. Let us look at a simple example in table 4, where the order is determined by the second letter of the internal form, where the sequence is अ उ र .

Table 4 - An example of sorting.

word	romanised	internal form	sort order
पुरिया	(puriya)	प उ र इ य आ	2nd
पति	(pati)	प अ त इ	1st
प्रयत्न	(prayatna)	प र अ य अ त न अ	3rd

3. Unicode tables and rendering rules.

The Unicode table follows very simply from the alphabet table 3, to which must be added the numerals. Punctuation symbols and mathematical symbols come elsewhere in the Unicode tables - strictly the numerals also come there, but we need the way that they are written to be able to be determined by Nepal usage, and not be forced to use the Indo-Arabic "International" set.

Table 5 Unicode Table.

	0x0	0x1	0x2	0x3	0x4	0x5	0x6	0x7
0	०	ॠ	अ	क	थ	ह	.	
1	१	.	आ	ख	द	क्ष	CM1	
2	२	:	इ	ग	ध	त्र	CM2	
3	३	ऽ	ई	घ	न	ज्ञ	CM3	
4	४	ॠ	उ	ङ	प	ॢ	CM4	
5	५	VM1	ऊ	च	फ			
6	६	VM2	ऋ	छ	ब			
7	७	VM3	ॠ	औ	भ			
8	८	VM4	लृ	ॡ	म			
9	९		ए	ज	य			
A			ऐ	ट	र		ॠ	
B			ओ	ठ	ल		ॡ	
C			औ	ड	व			
D			अं	ढ	श			
E			अः	ण	ष			
F		ॠ		त	स			

Sequences of these are rendered as shown in Table 6. Note that these are sample renderings, and the exact way sequences are rendered forms part of the font or style of writing. One method of rendering is to maintain a second table of glyphs, as seen in the Win31 table in Section 4, and then map sequences of codes to sequences of glyphs - see Part III for more details.

Table 6. Rendering

internal sequence	renderings
v ड क	अङ्क, अंक
v ॠ	अँ अँ ओँ ईँ

4. Bridging code table

This table must include all the important conjuncts and partial letters necessary, as well as matra forms of the vowels and other diacritics in enough variants to give a reasonable rendition of the languages of Nepal for the TTF font format. The following code table is proposed. If appropriately constructed it could be used as a glyph table for the rendering of the Unicode table, though in practice such glyph tables need not be restricted in size to just 234 glyphs.

Table 7. The code table for bridging use

	0	1	2	3	4	5	6	7	8	9
30				!	"	#	रु	%	&	'
40	()	*	+	,	-	.	/	०	१
50	२	३	४	५	६	७	८	९	:	;
60	<	=	>	?	@	ˆ	ˆ	ˆ	ˆ	ˆ
70	ˆ	ˆ	:	ॐ	अ	आ	इ	ई	उ	ऊ
80	ऋ	ॠ	लृ	ए	ऐ	ओ	औ	क	ख	क
90	क्र	[\]	ख	_	रु	ग	ग	घ
100	च	ड	ड्	ड्व	ङ	ञ	ट	ट्ट	ठ	छ
110	ज	ज	झ	ञ	ञ	ञ	ट	ट्ट	ठ	ठ
120	ठु	ड	डु	{		}	ढ		ण	ण
130	त	त	त	...	थ	थ	द	द	द्व	द्व
140	द्य	द्व		ध	न	'	'	"	"	="
150	-	—	प	™	र	फ	फ	ब	ब	भ
160		भ	म	म	म	य		र	च	©
170	र	«	»	८	रु	रु	०	ल	ल	व
180	ॠ	व	॥	·	श	श	श	»	¼	½
190	¾	श्र	श्र	ष	ठ	स	स	स	स	ह
200	ह	ह	ह	ह	ह	ह	ह	क्ष	क्ष	त्र
210	ज्ञ	ज्ञ	ऽ	।	ि	ि	ी	ी	ु	ु
220	ु	ु	ु	ु	ु	ु	ु	ु	ु	ु
230	ु	ु	ु	ु	ु	ु	ु	ु	ु	ु
240	ु	ु	ु	ु	ु	ु	ु	ु	ु	ु
250	ु	ु	ु	ु	ु	ु	ु	ु	ु	ु

5 Keyboards

Two keyboard layouts are defined - shown in Figures 1 and 2. The mappings from these keyboards to the internal code in either the Unicode or the bridging code is left to the implementors.

Figure 1. Remington Keyboard Layout

ज ~	ञ	द	घ	ड	ढ	ट	ठ	ड	ढ	ण	॰	+	
ज `	१	२	३	४	५	६	७	८	९	०	-	=	\
tab	त्र Q	ध W	भ E	च R	त T	थ Y	ग U	ङ I	झ O	ए P	{	^	enter
	त्र	ध	भ	च	त	थ	ग	ङ	झ	उ	[^	
cap lk	ब A	क S	म D	न F	ग G	ङ H	ज J	घ K	झ L	ः	“	’	
	ब	क	म	न	ग	ङ	ज	घ	झ	स ;	“	’	
shift	श Z	ह X	अ C	ख V	द B	ल N	: M	ड <	श्र >	रू ?	shift		
	श	ह	अ	ख	द	ल	ः	, ,	.	र /			
space													

Figure 2. Phonetic Keyboard

ज ~	!	@	#	रू Rs.	%	^	ऊँ	*	()	-	+	
ज `	१	२	३	४	५	६	७	८	९	०	-	=	\
tab	त्र Q	व W	ए E	र R	थ T	य Y	उ U	इ I	ओ O	फ P	{	ढ }	enter
	क्ष	व	ए	र	थ	य	उ	इ	ओ	फ	[ढ]	
cap lock	अ A	श S	ध D	् F	घ G	ह H	ज J	ख K	ल L	:	“	’	
	अ	श	ध	्	घ	ह	ज	ख	ल	;	“	’	
shift	अ Z	ष X	च C	ं V	भ B	ण N	ं M	ँ <	ड >	?	shift		
	अ	ष	च	ं	भ	ण	ं	, ,	.	/			
space													

Part III. Guidance for Implementers.

There are many items that must be co-ordinated when making a font for a computer whether it is a TTF (True Type Font) for Windows 3.1 and Windows 95 or if it is a Unicode font for Windows NT or Windows 98. First you must start with a language, identify the alphabet, and determine what all the characters are and how they are used to form words and sentences. Letters of the alphabet have to be typed into the computer from the keyboard, must show on the screen with the correct shapes, must be saved on a disk in a file for future use, and must be printed on paper with printers in the correct shape for other people to recognise and read. If a list of items is made then it is helpful if you can properly alphabetize the list in a sorted order so a word will be found in the list in the place that people would expect to find it. So where do you start to make a font? Most people like to start with the printed output, trying to make it look pleasing to the reader, but the most important place to start is with what is to be stored on the disk since that is the real 'data' that the computer is using to do it's work, and it is the part which needs to correctly represent the characters and sounds as they are used in the language. The rest of the process is either 'input' or 'output' or 'processing' the data to analyze it or rearrange it. The 'input' is usually from a keyboard and 'output' can be to a screen, printer, disk, or email.

The first step in making a new font is to decide on the basic building blocks of the language which is the alphabet which is listed in table 3. The list actually defines two things which is the alphabet characters and their shapes. It is important now to give each character a unique computer code so that the 'character code' can be used to save the data on a disk. This standard includes all of the alphabets of the languages in Nepal so there will be a few extra characters in the list as shown in Table 5 for Unicode fonts or Windows TTF fonts. This section will only deal with Unicode fonts so see the next section for issues related to TTF fonts. The reason for including TTF fonts in this standard is due to the fact that many people will be still using Windows 3.1 for a few years to come on older computers even though Unicode fonts will be supported on new systems.

All of the characters which are used in the language can be built from the list in table 5 by using combinations of the characters. It is important that the reader properly understands the idea of how the *explicit 'A'* (अ) affects document encoding and font creation. Normally when you type a consonant there is an *implicit 'अ'* unless a 'halant' is added to the character or another vowel is added. With Unicode fonts the proper way to designate a full consonant will be to add an *explicit 'अ'* after the character so a क becomes क अ instead of just क, while a के becomes क ए and a क्र becomes क र अ. Half characters like ऋ or ॠ can be represented with just a क and ण with no trailing अ. This system for representing characters on the disk's data is much more accurate than what is done in TTF fonts today and will maintain much better integrity when the data is sorted or analyzed.

Conjunct characters are made when two consonants are combined to make a new shape which is partly or totally different from the two characters that it is made from. Some examples are the dya द्य and sna स्न conjuncts where the dya द्य is made from 'd y a' द य अ and the sna स्न is made from 's n a' स ण अ (not sa na). The missing explicit 'a' अ from between the first two consonants indicates that they are to be joined together. In Unicode fonts it will be up to the designer's discretion as to how many conjuncts to make but it is recommended that at least the ones shown in this standard should be included in the font.

See table 8 for a complete list of character glyphs and the sequences used to represent them. Note that these sequences are what is saved on the disk and does not need to resemble what you type to input the data from the keyboard. That will be dealt with when considering the keyboard layout.

Table 8. The code table for most common glyphs

glyph	code	glyph	code	glyph	code	glyph	code	glyph	code
अ	अ	ख	खअ	त	तअ	र	र	ह	हअ
आ	आ	ङ	ङअ	थ	थअ	र	र	ह	हअ
इ	इ	घ	घअ	थ	थ	र	र	ह	हअ
ई	ई	ह	हअ	द	दअ	र	र	क्ष	क्षअ
उ	उ	च	चअ	द	दअ	रु	रु	क्ष	क्ष
ऊ	ऊ	च	च	द	दअ	रु	रु	त्र	त्रअ
ऋ	ऋ	छ	छअ	द	दअ	ल	लअ	ज्ञ	ज्ञअ
ॠ	ॠ	ज	जअ	द	दअ	ल	ल	ज्ञ	ज्ञ
लृ	लृ	ज	ज	द	दअ	व	वअ	।	।
लृ	लृ	ञ	ञअ	ध	धअ	व	व	।	।
ए	ए	रु	रुअ	ध	ध	व	वअ	।	।
ऐ	ऐ	रु	रु	न	नअ	श	शअ	।	।
ओ	ओ	ज	जअ	न	न	श	शअ	।	।
औ	औ	ञ	ञ	प	पअ	श	श	।	।
अं	अं	ट	टअ	प	प	श	शअ	।	।
अः	अः	ट	टअ	फ	फअ	श	शअ	।	।
क	कअ	ठ	ठअ	फ	फ	श	शअ	।	।
क	क	ठ	ठअ	फ	फअ	ष	षअ	।	।
क्त	क्तअ	ठ	ठअ	ब	बअ	ष	ष	।	।
क्र	करअ	ड	डअ	ब	ब	स	सअ	।	।
ख	खअ	ड	डअ	भ	भअ	स	स	।	।
रु	ख	ड	डअ	भ	भ	स	सनअ	।	।
ग	गअ	ढ	ढअ	म	मअ	स	सरअ	।	।
ग	ग	ढ	ढअ	म	म	ह	हअ	।	।
घ	घअ	ण	णअ	य	यअ	ह	हनअ	।	।
घ	घ	ण	ण	य	य	ह	हअ	।	।
ङ	ङअ	त	तअ	य	यअ	ह	हअ	।	।
ङ	ङअ	त	त	र	रअ	ह	हरअ	।	।

Now the fun part comes of putting some of this together and trying to understand it. To type a conjunct character you enter the sequences for the parts and the system will show the conjunct. Assigning a certain 'character code' sequence to a 'glyph' or character image is done in the program used to create the characters so check the manuals for creating Unicode Fonts. In Windows TTF fonts the keyboard mapping program has to be programmed to substitute a different character depending on the typing sequence entered so check the section below about TTF Fonts.

Drawing the Character Shapes 'Glyphs'

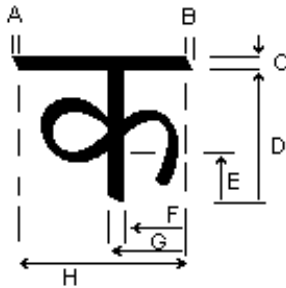
To make a font distinct it must be given a unique 'Name' and it must have consistency in the style that the characters are drawn with. Some fonts use plain block shapes while others may be slanted like *italics* or look like they were drawn with a 'flat' pen so that on a letter like 'O' two sides will be thick and two sides will be thin. What style the font is drawn in must be decided before drawing even the first character so that they will all be balanced in shape and style. Then it is important to pick a few standard measurements. The base line is usually taken as the bottom of the consonant character with the 'Ascent' height being to the top of the '।' character and the 'Descent' height being to the bottom of the '।' vowel. The top 'bar' which connects characters should be at about 2/3 of the 'Ascent' figure. It

is also important to decide on basic widths for characters in reference to the vertical stem even though some characters will not have a stem. For standard TTF fonts these numbers in Fontographer may be:

Ascent height : 2000 Descent height : -500 Bar height : 1300
 Minimum width : 700 Maximum width : 2000 Underline height : -300
 Right offset for right aligned : 200 Right offset for center aligned : 600

With these measurements the 'क' character would be 1300 tall and around 1600 wide with the vertical stem at 600 from the right 'width' line or at 1000 from the left edge. This means the base of the ' ँ ' would start 600 left of the zero line for center aligned and 200 left of the zero line for right aligned characters. The need for two sets of diacritics is for this reason that some characters have the vertical stem in the middle and some have it on the left edge. A list is given below of special characters which need to be drawn with a specific measurement in mind.

This drawing shows the measurements which need to be planned before starting to draw the font characters. Use the table to plan it carefully. The 'Origin' of the character is the bottom left corner so 'H' measures to the right from the Origin and 'D' measures up from the origin.



Label	Center aligned character क	Right aligned character ऀ	Description
A	60	60	slant on top bar can be square or slanted
B	60	60	slant on top bar - same as A
C	130	130	top bar thickness
C+D	1300	1300	C+D : is height of character
E	650	650	height of half 'r' ↘
F	700	300	diacritics align on F and G measurements
G	830	430	alignment edge for diacritics and vertical stem. This is one of the most important
G-F	130	130	width of vertical stem
H	1600	900	width of character - top bar should extend 5 points beyond this measurement

Note: all numbers are given in co ordinates of : (horizontal), (vertical) where the bottom left corner is taken as 0,0
 There are a number of groups of characters which need to be dealt with consistently in all fonts so that the characters and diacritics will align properly. Otherwise if one font wants the ऀ to be center aligned and another wants it right aligned then changing fonts will cause the diacritics to be misaligned and there will again be a problem. This problem will only be for the TTF fonts since in Unicode there is the possibility to program into the font how to align the diacritics over the consonants. Here are the sets of characters to plan for:

Table 8 Sets of characters

Characters which have half forms	क ख ग घ च ज ङ ञ त थ ध न प फ ब भ म य र ल व श ष स क्ष ज्ञ
Center aligned characters	ऋ ॠ क क्त ऋ ङ् फ रुरु
Characters for tent 'r' ँ	ड छ ट ठ ड ढ
Characters for lowered 'u' उ	कृ क्ख क्ख ङ्ङ टु टु डु डु
Characters for extended reach 'ikar' ि	ख ग घ च छ ज थ ध ल श स
Characters for extended reach 'iikar' िी	क क्त ऋ ङ् फ

Data Entry

Data entry is normally done from the keyboard and in Nepal this has meant a close copy of the Devanagiri Remington Typewriter's keyboard but in this standard it is possible to offer many different keyboard layouts which the typist will pick from at the time of entering or editing data. The data being stored on the disk will always be the same so it does not matter which keyboard is used to enter or edit the data. When installing the fonts on a computer it will be necessary to also install the KEYMAN program to control the keyboard layouts so installation instructions must be given to explain this. The Keyman program is written by Marc Durdin and distributed by Tavultesoft freely.

Windows 3.1 & 95 TTF font creation

The above discussions apply quite well to the new standard for TTF fonts except that the options for 'glyphs' which are represented by multiple 'character codes' is not possible so a more complete list of characters must be assigned directly to codes and all of the work of translating typing sequences must be interpreted by the keyboard program. For this part of the standard refer to the ASCII table 7 which shows the character code assignments and a program is provided called KEYMAN which has to be preprogrammed to correctly choose characters, half characters, and conjuncts according to the standard set out for Unicode fonts. There are some characters missing due to lack of space but be careful not to use the spaces which do not have Nepali characters since many of them are used by programs for their own formatting purposes.

There are a number of other things to be noted in this standard and that is the inclusion of double sets of diacritics. This is to allow for proper alignment of diacritics on those consonants which have the vertical stem in the center of characters like क instead of on the right side as with the ग. Otherwise an

alignment problem occurs as seen here. के गे के गे The first two use the left aligned `

diacritic while the second two use the center aligned ` diacritic. In the TTF table 7 above, the first of the pair is for right alignment and the second is for center alignment. So for the ` the right aligned one is at decimal 232 while the right aligned one is at decimal 233. The characters which need the center aligned diacritics are listed in table 8 above. All other characters if they need a diacritic will use the right aligned diacritic. Extra characters in the font that need to be noted are as follows.

- d69 extra ` for use with right aligned 'ekar' `
- d220 lowered ु for tall conjuncts
- d225 lowered ू for tall conjuncts
- d234 extra ` for alternat Limbu form (ie `)

d242	dot below for right aligned and ङ ञ
d243	dot below for center aligned
d244	dot below for व ज फ
d245	dot in between middle of character
d246	dot in line with top bar line - Tibetan fonts
d250	glottal stop
d251,252	Sanskrit character "gu"
d254	symbol for showing relation of diacritic to character
d255	thin space

There are two keyboard mappings provided with the font distribution which are named 'NEP_PHON.KMN' and 'NEP_TYPE.KMN' which have comments in them for how to understand or edit them. The font distribution is available from <http://www.nepaliug.org.np>.