Collation of Myanmar (Burmese) in Unicode

Sorting Myanmar in Unicode according to "Spelling Book Order"

Introduction

This document presents an algorithm for sorting text in the Myanmar language, which is sometimes still referred to as Burmese. There are two main sorting orders that have been used with Myanmar: "Pali Order" and "Spelling Book Order". The former was used in older dictionaries, whereas variations on the latter are used in most modern dictionaries.[1]

The algorithm presented here will focus on the "Spelling Book Order" as it is used in modern Myanmar. There are some subtle variations to this order used by different dictionaries, but the order used here will be based on the Myanmar Language Commission Spelling Dictionary.[2]

There are many different non-Unicode encodings for Myanmar text, however, these are based on glyphs rather than linguistic symbols. It is possible to type the same word in different ways, using the same encoding, whilst keeping the spelling the same. For example, အမှူး or အမှူး (leader), both are spelled in the same way but using a variation on the glyph for the uu vowel (typed as *trSL;* and *tr*^{a}*;* respectively in the WinInnwa font). This makes collation hard because lots of combinations have to be worked out.

Collation in Unicode is simpler because in most cases there is no variation in how the word is spelled in terms of code points. The choice of which glyphs to use is made by the font, not the typist. This document will assume that Myanmar is encoded according to Unicode 5.1.0[3] and Unicode Technical Note 11 (revision 2).[4]

The notation used here is intended for the purposes of collation only and sometimes may not represent the normal linguistic nomenclature.

Collation Elements

Myanmar is collated based on syllables. A Myanmar syllable encoded in Unicode can be broken into 5 parts for collation:

<consonant><medial><vowel><final><tone>

Only the consonant is always present, one or more of the other parts may be empty in any given syllable. In practice the vowel may be displayed before the consonant e.g. ကေ, but it is encoded as U+1000 (Myanmar letter KA က) U+1031 (Myanmar vowel sign E ေ).

The syllable needs to be reordered for collation, because the final has a higher priority than the vowel:
<consonant>, <medial>, <final>, <vowel>, <tone>.

[1] *Burmese: An Introduction to the Script*, John Okell, 1994, SOAS, Appendix 6: Alphabetic Order.

[2] Myanmar Spelling Dictionary, Myanmar Language Commission, Second Edtion, 2003.

[3] The Unicode Consortium. The Unicode Standard, Version 5.1.0, defined by: *The Unicode Standard, Version 5.0* (Boston, MA, Addison-Wesley, 2007. ISBN 0-321-48091-0), as amended by Unicode 5.1.0 (http://www.unicode.org/versions/Unicode5.1.0/).

[4] Unicode Technical Note 11: *Representing Myanmar in Unicode: Details and Examples*, Martin Hosken & Maung Tuntunlwin, 2007-01-25, http://www.unicode.org/notes/tn11/ .

Each of these parts of the syllable may be composed of one or more characters as the following tables show.

Consonant

Collation Order – read left to right and then down. The data is presented in the traditional layout of the alphabet.

Glyph		Code	Glyph		Code	Glyph		Code	Glyph		Code	Glyph		Code	Glyph		Code
C1	က	U+1000	C2	ခ	U+1001	C3	ဂ	U+1002	C4	ဃ	U+1003	C5	င	U+1004			
C6	စ	U+1005	C7	ဆ	U+1006	C8	ဇ	U+1007	C9	ဈ	U+1008	C9	ဉ	U+1009	C10	ည	U+100A
C11	ဋ	U+100B	C12	ဌ	U+100C	C13	ဍ	U+100D	C14	ဎ	U+100E	C15	ဏ	U+100F			
C16	တ	U+1010	C17	ထ	U+1011	C18	ဒ	U+1012	C19	ဓ	U+1013	C20	န	U+1014			
C21	ပ	U+1015	C22	ဖ	U+1016	C23	ဗ	U+1017	C24	ဘ	U+1018	C25	မ	U+1019			
C26	ယ	U+101A	C27	ရ	U+101B	C28	လ	U+101C	C29	ဝ	U+101D	C30	သ	U+101E			
			C31	ဟ	U+101F	C32	ဠ	U+1020	C33	အ	U+1021						

Note 1: The relative order is the same as the code points themselves. However, if the collation is extended to the additional code points for Karen, Mon and Shan, then it will become significantly more complicated, with several more consonants inserted.

Note 2: C33 is actually the A vowel, but it behaves in many ways like a consonant in regards to the other parts of the syllable.

Note 3: It may be appropriate to append the “Various Signs” U+104C ... U+104F at the end of this class – see comments in *Other Myanmar Characters* below.

Medials

The case where there is no medial is also included so that the relative sequence is clear.

Order	*Glyph(s)*	*Unicode Sequence*
M0	–	-
M1	◌ျ	U+103B
M2	◌ြ	U+103C
M3	◌ွ	U+103D
M4	◌ှ	U+103E
M5	◌ျွ	U+103B U+103D

Order	*Glyph(s)*	*Unicode Sequence*
M6	◌ြွ	U+103C U+103D
M7	◌ျှ	U+103B U+103E
M8	◌ြှ	U+103C U+103E
M9	◌ွှ	U+103D U+103E
M10	◌ျွှ	U+103C U+103D U+103E
M11	◌ြွှ	U+103B U+103D U+103E

Note 4: the combined medials are treated as one unit for collation not as a sequence of component medials.

Vowels

Collation order for Vowels – read down, then left to right.

Order	*Glyph*	*Unicode Sequence*	*Order*	*Glyph(s)*	*Unicode Sequence*
V0	-	-	V6	ေ◌	U+1031
V1	◌ါ/◌ာ	U+102B/U+102C	V7	◌ဲ	U+1032
V2	◌ိ	U+102D	V8	ေ◌ါ/ ေ◌ာ	U+1031 U+102B / U+1031 U+102C
V3	◌ီ	U+102E	V9	ေ◌ါ်/ ေ◌ာ်	U+1031 U+102B U+103A / U+1031 U+102C U+103A
V4	◌ု/ ု	U+102F	V10	◌ံ	U+1036 – see note below
V5	◌ူ/ ူ	U+1030	V11	◌ို	U+102D U+102F

Note 5: V9 is actually the low tone form of V8, but it is included here, because it does not use the normal tone marks. There is never a final after V3, V5, V9 or V10.

Note 6: V10 is not really a vowel, however, when there is no other vowel it is treated as one for collation. When it occurs in the sequence U+102D U+1036 or U+102F U+1036, then U+1036 is collated as if it was a final U+1019 U+103A, which is what it is linguistically. It is slightly more complicated than this – see below for the details.[5]

Finals

Finals are marked with a Myanmar Sign Asat U+103A or with Myanmar Sign Virama U+1039 when the consonant of the following syllable is to be displayed underneath the final. In a few rare cases, a ligature of the final and the following consonant is used instead, but these are rendering issues and are not relevant for collation.

[5] This is in contrast to John Okell, who collates U+1036 as equivalent to U+1019 U+103A. e.g. *Burmese/Myanmar Dictionary of Grammatical Forms*, John Okell & Anna Allott, 2001, Curzon Press.

Glyph		Code	Glyph		Code	Glyph		Code	Glyph		Code	Glyph		Code	Glyph		Code
F0	-	-	F1	က်	U+1000 U+1039/ U+1000 U+103A	F2	ခ်	U+1001 U+1039/ U+1001 U+103A	F3	ဂ်	U+1002 U+1039/ U+1002 U+103A	F4	ဃ်	U+1003 U+1039/ U+1003 U+103A	F5	င်	U+1004 U+103A U+1039/ U+1004 U+103A
F6	စ်	U+1005 U+1039/ U+1005 U+103A	F7	ဆ်	U+1006 U+1039/ U+1000 U+103A	F8	ဇ်	U+1007 U+1039/ U+1007 U+103A	F9	ဈ်	U+1008 U+1039/ U+1008 U+103A	F10	ဉ်	U+1009 U+1039/ U+1009 U+103A	F11	ည်	U+100A U+1039/ U+100A U+103A
F11	ဋ်	U+100B U+1039/ U+100B U+103A	F12	ဌ်	U+100C U+1039/ U+100C U+103A	F13	ဍ်	U+100D U+1039/ U+100D U+103A	F14	ဎ်	U+100E U+1039/ U+100E U+103A	F15	ဏ်	U+100F U+1039/ U+100F U+103A			
F16	တ်	U+1010 U+1039/ U+1010 U+103A	F17	ထ်	U+1011 U+1039/ U+1011 U+103A	F18	ဒ်	U+1012 U+1039/ U+1012 U+103A	F19	ဓ်	U+1013 U+1039/ U+1013 U+103A	F20	န်	U+1014 U+1039/ U+1014 U+103A			
F21	ပ်	U+1015 U+1039/ U+1015 U+103A	F22	ဖ်	U+1016 U+1039/ U+1016 U+103A	F23	ဗ်	U+1017 U+1039/ U+1017 U+103A	F24	ဘ်	U+1018 U+1039/ U+1018 U+103A	F25	မ်	U+1019 U+1039/ U+1019 U+103A			
F26	ယ်	U+101A U+1039/ U+101A U+103A				F28	လ်	U+101F U+1039/ U+101F U+103A				F30	သ်	U+101E U+1039/ U+101E U+103A			
						F32	ဠ်	U+1020 U+1039/ U+1020 U+103A									

Note 7: ဿ U+103F is a special case and is collated as if it was U+101E U+1039 U+101E.

Note 8: Kinzi is encoded as U+1004 U+103A U+1039. U+1004 U+1039 on its own does not exist in modern day Myanmar.

The difference between a normal final or killed consonant (KC) with U+103A and a stacked consonant (SC) with U+1039 can be treated at a secondary level.

Tones

Order	*Glyph*	*Unicode*
T0	–	-
T1	◌့	U+1037
T2	◌း	U+1038
T3	◌့း	U+1037 U+1038

Note 9: T3 is not normally found in a dictionary, it marks a genitive in some situations. It might however be found in book indexes etc.

Independent Vowels

The Independent vowels are collated as if they were written with အ U+1021 (Myanmar letter A) and the corresponding vowel. In some cases they may be followed by း U+1038 (Myanmar sign visarga), which collates in the same way as normal.

Order	*Glyph*	*Unicode Sequence*	*Equivalent Representation*	*Equivalent Sequence for Collation*	*Equivalent Collation Elements*
IV1	ဣ	U+1023	အိ	U+1021 U+102D	C33 V2
IV2	ဤ	U+1024	အီ	U+1021 U+102E	C33 V3
IV3	ဥ	U+1025	အု	U+1021 U+102F	C33 V4
IV4	ဦ	U+1026 or (U+1025 U+102E)	အူ	U+1021 U+1030	C33 V5
IV5	ဧ	U+1027	အေ	U+1021 U+1031	C33 V6
IV6	ဩ	U+1029	အော	U+1021 U+1031 U+102C	C33 V8
IV7	ဪ	U+102A	အော်	U+1021 U+1031 U+102C U+103A	C33 V9

Note 10: although these are equivalent for the purposes of collation, usually only one representation is correct in a given word.

Note 11: the independent vowels may take finals, tones and IV3 takes V4 and V10 as shown below.

The known independent vowel / final combinations are:

Independent Vowel / Final Combination	*Glyphs*	*Collation Elements*

`U+1025 U+1000 U+1039 /`
`U+1025 U+1000 U+103A`
	ဥက်	C33 F1 V4

`U+1023 U+1005 U+1039 /`
`U+1023 U+1005 U+103A`
	ဣစ်	C33 F6 V2

`U+1025 U+1005 U+1039 /`
`U+1025 U+1005 U+103A`
	ဥစ်	C33 F6 V4

`U+1027 U+100A U+1039 /`
`U+1027 U+100A U+103A`
	ဧည်	C33 F11 V6

`U+1023 U+100B U+1039 /`
`U+1023 U+100B U+103A`
	ဣဋ်	C33 F12 V2

`U+1025 U+100F U+1039 /`
`U+1025 U+100F U+103A`
	ဥဏ်	C33 F15 V4

`U+1023 U+1010 U+1039 /`
`U+1023 U+1010 U+103A`
	ဣတ်	C33 F16 V2

`U+1025 U+1010 U+1039 /`
`U+1025 U+1010 U+103A`
	ဥတ်	C33 F16 V4

`U+1023 U+1012 U+1039 /`
`U+1023 U+1012 U+103A`
	ဣဒ်	C33 F18 V2

`U+1023 U+1014 U+1039 /`
`U+1023 U+1014 U+103A`
	ဣန်	C33 F20 V2

`U+1025 U+1015 U+1039 /`
`U+1025 U+1015 U+103A`
	ဥပ်	C33 F21 V4

`U+1023 U+1019 U+1039 /`
`U+1023 U+1019 U+103A`
	ဣမ်	C33 F25 V2

`U+1025 U+102F U+1036`	ဥုံ	C33 F25 V4

`U+1023 U+101E U+1039 /`
`U+1023 U+101E U+103A`
	ဣသ်	C33 F30 V2

Independent Vowel / Final Combination	*Glyphs*	*Collation Elements*

`U+1025 U+101E U+1039 /`
`U+1025 U+101E U+103A`
	ဥသ်	C33 F30 V4

Note 12: The combination of U+1025 (Myanmar Letter U) with U+1036 (Myanmar Sign Anusvara) has the U+102F explicitly encoded and displayed. This situation is actually more complicated as described below.

The difference between a normal vowel (NV) and the independent vowel (IV) form can be treated at the tertiary level.

Anusvara and Matha

The difference between Anusvara U+1036 and U+1014 U+103A needs to be taken at a primary level when it occurs after the vowels U+102D and U+102F. However, it is not behaving like a completely different killed consonant because occurrences with U+1037 and U+1038 also need to be included as illustrated in the table below.[6] In this case, the vowel, tone and anusvara effectively become mixed up into a combined vowel key.

Encoding	*Glyphs*	*Primary Key*	*Tertiary Key*
`U+1021 U+102F U+1019 U+103A`	အုမ်	C33 F25 V4.1	- - NV
`U+1025 U+1019 U+103A`	ဥမ်	C33 F25 V4.1	- - IV
`U+1021 U+102F U+1036`	အုံ	C33 F25 V4.2	- - NV
`U+1025 U+102F U+1036`	ဥုံ	C33 F25 V4.2	- - IV
`U+1021 U+102F U+1019 U+103A U+1037`	အုမ့်	C33 F25 V4.3	- - NV
`U+1021 U+102F U+1036 U+1037`	အုံ့	C33 F25 V4.3	- - NV
`U+1021 U+102F U+1019 U+103A U+1038`	အုမ်း	C33 F25 V4.4	- - NV
`U+1021 U+102F U+1036 U+1038`	အုံး	C33 F25 V4.4	- - NV

The independent vowel examples may require special handling to allow for the consonant and vowel being together for encoding, but apart for collation.

Contractions

There are a few words which are written with a repeated consonant omitted, but which should be collated as if the consonant was still present. The number of these is small, though the list below is probably not complete.

[6] This is following the order given in the Myanmar Language Commission Myanmar-English Dictionary, 2006.

Word	***Meaning***	***Unicode Representation***	***Collation Equivalent***	***Collation Elements***
ယောက်ျား	man	U+101A U+1031 U+102C U+1000 U+103A U+103B U+102C U+1038	U+101A U+1031 U+102C U+1000 U+103A **U+1000** U+103B U+102C U+1038	C26 V8 F1 C1 M1 V1 T2
ကျွန်ုပ်	1st person singular	U+1000 U+103B U+103D U+1014 U+103A U+102F U+1015 U+103A	U+1000 U+103A U+103D U+1014 U+103A **U+1014** U+102F U+1015 U+103A	C1 M5 F20 C20 V4 F21

Short Forms

These are variations on the normal spelling, that are still found in current use. If they need to be collated, then they should be collated with their normal spelling not the variant. However, implementing this level of collation support for Myanmar can probably be regarded as optional. They can be thought of as “short forms” because their left to right width is less.

Word	***Short Form***	***Normal Spelling***	***Short form Unicode Sequence***	***Normal Unicode Sequence***	***Collation Elements***
daughter	သ္မီး	သမီး	U+101E **U+1039** U+1019 U+102E U+1038	U+101E U+1019 U+102E U+1038	C30 C25 V3 T2
cooked rice	ထ္မင်း	ထမင်း	U+1011 **U+1039** U+1019 U+1004 U+103A U+1038	U+1011 U+1019 U+1004 U+103A U+1038	C17 C25 F5 T2
tea	လ္ဘက်	လက် ဖက် or လက် ဘက်	U+101C U+1039 U+1018 U+1000 U+103A	U+101C **U+1000** U+103A *U+1018* U+1000 U+103A	C28 F1 C25 F1
right hand side	လကျ်ာ	လက်ယာ	U+101C U+1000 U+103A U+103B U+102C	U+101C U+1000 U+103A U+101A U+102C	C28 F1 C26 V1

Note 13: In the first 2 examples the U+1039 should be ignored and is purely a trick to get the correct rendering.

Note 14: The third example is different in that the final က် has been dropped from the first consonant, so the U+1000 and U+103A have been removed and replaced with U+1039. The second consonant is normally now spelled as U+1016 ဖ (PHA) not U+1018 ဘ (BHA), but it should probably still be collated as U+1018.

Other Myanmar Characters

The Myanmar symbols may be collated as if spelt with their full Burmese names:

Symbol	***Symbol Encoding***	***Name***	***Equivalent Unicode Sequence***
၌	U+104C	နှိုက်	U+1014 U+103E U+102D U+102F U+1000 U+103A
၍	U+104D	ရွေ့	U+101B U+103D U+1031 U+1037
၎င်း	U+104E U+1004 U+103A U+1038	လည်းကောင်း	U+101C U+100A U+103A U+1038 U+1000 U+1031 U+102C U+1004 U+103A U+1038
၏	U+104F	အိ	U+1021 U+102D

၊ Myanmar sign little section (U+104A) and ။ Myanmar sign section (U+104B) can normally be ignored or treated at a lower level similar to collation of punctuation in other languages.

Myanmar digits can be treated at a primary level as equal to the digits of other languages and at a secondary level on a script basis as per the Unicode standard.

Implementation

The <vowel><final> needs to be combined, because syllables with both are collated first by the final, then by the vowel. The <tone> part can be treated separately.

Special care needs to be taken to handle U+102D U+1036, U+102F U+1036 including the case where it is preceded by U+103F. The cases where an independent vowel and final are combined may need to be explicitly listed to give the correct results.

The relative precedence of the 5 syllable parts is:

1. <consonant> C
2. <tone> T
3. <vowel> V
4. <final> F
5. <medial> M

The following combinations, given in encoding order, are permissible and illustrate this precedence:

1. C
2. CV
3. CVT
4. CF
5. CFT
6. CVF (CFV)
7. CVFT (CFVT)
8. CM
9. CMV
10. CMVT
11. CMF
12. CMFT
13. CMVF (CMFV)
14. CMVFT (CMFVT)

The relative order of the <tone> and <vowel> is only important if the <vowel> and <final> are reordered explicitly before collation (as shown in brackets) rather than being combined into a single collation unit. If such reordering is done, then the number of collation units drops dramatically.

Glibc

An implementation has been written for Glibc. This combines <vowel><final> as one collation unit to allow the <final> to take precedence over the vowel. This gives a large number of collation elements, but gives correct results. Performance is probably sub-optimal because of the large number of collation elements used.

ICU

An implementation has also been written for ICU. It uses a large number of collation elements, combining <vowel><final> to ensure the correct sequence. There is probably a lot of scope to optimise it, but it might require changes to the ICU source code.

Examples

The table below shows a selection of words to illustrate the 5 different orders of collation.

Word	*Collation Element in 1^{st} Syllable*					*Unicode Sequence*
	Consonant	*Medial*	*Final*	*Vowel*	*Tone*	
ကခုန်	C1	M0	F0	V0	T0	**U+1000** U+1001 U+102F U+1014 U+103A
ကာ	C1	M0	F0	V1	T0	**U+1000 U+102C**
ကား	C1	M0	F0	V1	T2	**U+1000 U+102C U+1038**
ကိရိယာ	C1	M0	F0	V2	T0	**U+1000 U+102D** U+101B U+102D U+101A U+102C
ကုဗပေ	C1	M0	F0	V4	T0	**U+1000 U+102F** U+1017 U+1015 U+1031
ကေဒါ	C1	M0	F0	V6	T0	**U+1000 U+1031** U+1012 U+102B
ကဲလွန်	C1	M0	F0	V7	T0	**U+1000 U+1032** U+101C U+103D U+1014 U+103A
ကဲ့	C1	M0	F0	V7	T1	**U+1000 U+1032 U+1037**
ကောလီကြေ	C1	M0	F0	V8	T0	**U+1000 U+1031 U+102C** U+101C U+102E U+1000 U+103C U+1031
ကော့လန်	C1	M0	F0	V8	T1	**U+1000 U+1031 U+102C U+1037** U+101C U+1014 U+103A
ကော်လံ	C1	M0	F0	V9	T0	**U+1000 U+1031 U+102C U+103A** U+101C U+1036
ကံ	C1	M0	F0	V10	T0	**U+1000 U+1036**
ကို	C1	M0	F0	V11	T0	**U+1000 U+102D U+102F**
ကက္ကရာ	C1	M0	F1	V0	T0	**U+1000 U+1000 U+1039** U+1000 U+101B U+102C
ကက်ကင်းဓာတ်	C1	M0	F1	V0	T0	**U+1000 U+1000 U+103A** U+1000 U+1004 U+103A U+1038 U+1013 U+102C U+1010 U+103A

Word	*Collation Element in 1st Syllable*					*Unicode Sequence*
	Consonant	*Medial*	*Final*	*Vowel*	*Tone*	
ကုက္ကလံ	C1	M0	F1	V4	T0	**U+1000 U+102F U+1000 U+1039** U+1000 U+101C U+1036
ကောက်ခံ	C1	M0	F1	V8	T0	**U+1000 U+1031 U+102C U+1000 U+103A** U+1001 U+1036
ကိုက်	C1	M0	F1	V11	T0	**U+1000 U+102D U+102F U+1000 U+103A**
ကင်မရာ	C1	M0	F5	V0	T0	**U+1000 U+1004 U+103A** U+1019 U+101B U+102C
ကင်းစီး	C1	M0	F5	V0	T2	**U+1000 U+1004 U+103A U+1038** U+1005 U+102E U+1038
ကုမ္ပဏီ	C1	M0	F25	V4	T0	**U+1000 U+102F U+1019 U+1039** U+1015 U+100F U+102E
ကုံင	C1	M0	F25	V4	T0	**U+1000 U+102F U+1036** U+1004
ကုံး	C1	M0	F25	V4	T2	**U+1000 U+102F U+1036 U+1038**
ကယ်ချွတ်	C1	M0	F26	V0	T0	**U+1000 U+101A U+103A** U+1001 U+103B U+103D U+1010 U+103A
ကျ	C1	M1	F0	V0	T0	**U+1000 U+103B**
ကျာ	C1	M1	F0	V1	T0	**U+1000 U+103B U+102C**
ကြ	C1	M2	F0	V0	T0	**U+1000 U+103C**
ကြောင့်	C1	M2	F4	V8	T1	**U+1000 U+103C U+1031 U+102C U+1004 U+103A U+1037**
ကွာခြား	C1	M3	F0	V1	T2	**U+1000 U+103D U+102C** U+1001 U+103C U+102C U+1038
ကျွေး	C1	M5	F0	V6	T2	**U+1000 U+103B U+103D U+1031 U+1038**
ကြွားဝါ	C1	M6	F0	V1	T2	**U+1000 U+103C U+103D U+102C U+1038** U+101D U+102B
ခမျာ	C2	M0	F0	V0	T0	**U+1001** U+1019 U+103B U+102C
အိတ်ကပ်	C33	M0	F16	V2	T0	**U+1021 U+102D U+1010 U+103A** U+1000 U+1015 U+103A
ဣတ္ထိလိင်	C33	M0	F16	V2	T0	**U+1023 U+1010 U+1039** U+1011 U+102D U+101C U+102D U+1004 U+103A
အုတ်	C33	M0	F16	V4	T0	**U+1021 U+102F U+1010 U+103A**

Note 15: Although several of the words are multi-syllable, the first syllable (code points in bold) is sufficient to determine the sort order in all but one case in this example. (The one exception is for the 2 examples with C1 M0 F1 V0 T0, where the second syllable controls sorting).

Conclusions

An algorithm for Myanmar Collation has been presented in terms of 5 collation elements within a syllable. In order of precedence these are: <consonant>, <tone>, <vowel>, <final>, <medial>, but they occur in the syllable in the sequence <consonant>, <medial>, <vowel>, <final>, <tone> and the <vowel> and <final> must be reordered for collation. The independent vowels should be collated as equivalent to the same vowel sound written with Myanmar letter A (U+1021). The collation is complicated because some code points may represent more than one of the syllable components. In addition, a complete implementation should handle the special cases for anusvara, contractions and short forms.

Last Updated: 24 April 2008
Keith Stribley