

PR 96: Background information for Malayalam

L2/07-173

Cibu C Johny
cibu@google.com
2007-May-13

Tighest context for preserving ZWNJ in Malayalam

[PR 96](#) is trying to find the context in which ZWNJ should be preserved. Following table is to aid that decision making, by showing tightest possible context for Malayalam.

There exists an entry in the table below, for row R and column C, iff, there exists a conjunct form for <R, VIRAMA, C> in Malayalam writing system. That means, if there exists an entry for row R and column C, deleting ZWNJ in the context of <R, VIRAMA, ZWNJ, C> might turn the form with visible virama to a conjunct without visible virama. This is risky, because the word contexts in which <R, VIRAMA, ZWNJ, C> appear might require explicit virama as in the case of 'ദ്രക്സാക്ഷ'(/drksaakssi/).

Similarly, if there is no entry for an R and a C, then ZWNJ in this context can be removed safely because no conjunct is there in the writing system for <R, VIRAMA, C>.

Simple regular expression tighter than <Consonant, Virama, ZWNJ, Consonant>

Pink rows X = {ക, ഗ, ച, ജ, ഞ, ഡ, ണ, ത, ദ, ന, പ, ഫ, ബ, യ, ഷ, സ, ഴ} and **purple** columns Y = {ക, ങ, ട, ത, ന, മ, യ, ര, ല, വ, ഉ, റ} together cover all entries in the table. So, the patterns <X, VIRAMA, ZWNJ, C> and <C, VIRAMA, ZWNJ, Y> are simple regular expressions to cover all ZWNJ preserving contexts for Malayalam.

U+0D..	ക	ഖ	ഗ	ഘ	ങ	ച	ഛ	ജ	ഝ	ഞ	ട	ഠ	ഡ	ഢ	ണ	ത	ഥ	ദ	ധ	ന	പ	ഫ	ബ	ഭ	മ	യ	ര	ല	വ	ശ	ഷ	സ	ഹ	ള	ഴ	റ		
ക (15)	ക										കു				കി	കത				കു					കു	കു	കി	കു	ക		ക്ഷ	കി					കി	
ഖ(16)																										ഖു	ഖ	ഖ	ഖ									
ഗ(17)			ഗ	ഗ															ഗ	ഗ					ഗ	ഗ	ഗ	ഗ										
ഘ(18)																				ഘ						ഘ	ഘ	ഘ	ഘ									
ങ(19)	ങ				ങ																						ങ	ങ	ങ	ങ								
ച(1A)						ച	ച																				ച	ച	ച	ച								
ഛ(1B)																											ഛ	ഛ	ഛ	ഛ								
ജ(1C)								ജ		ജ																	ജ	ജ	ജ	ജ								
ഝ(1D)																											ഝ	ഝ	ഝ	ഝ								
ഞ(1E)						ഞ	ഞ	ഞ		ഞ																	ഞ	ഞ	ഞ	ഞ								

s(1F)										ട										ട്യ	ട	ട്	ടാ							
o(20)																				ഓ	ഓ	ഓ	ഓ							
u(21)											യ്യ	യ്യ								യ്യ	യ്യ	യ്യ	യ്യ							
u(22)																				യ്യ	യ്യ	യ്യ	യ്യ							
ൺ(23)										ൺ	ൺ	ൺ	ൺ	ൺ						ൺ	ൺ	ൺ	ൺ	ൺ						
ത(24)														ത	ത					ത	ത	ത	ത			ത				
ഥ(25)														ഥ						ഥ	ഥ	ഥ	ഥ	ഥ						
ദ(26)															ദ	ദ				ദ	ദ	ദ	ദ							
ധ(27)														ധ						ധ	ധ	ധ	ധ	ധ						
ന(28)							ന							ന	ന	ന	ന			ന	ന	ന	ന	ന						
പ(2A)										പ				പ			പ	പ	പ		പ	പ	പ	പ			പ			
ഫ(2B)	ഫ									ഫ				ഫ	ഫ		ഫ	ഫ		ഫ	ഫ	ഫ	ഫ	ഫ			ഫ			ഫ
ബ(2C)			ബ				ബ							ബ	ബ	ബ			ബ	ബ	ബ	ബ	ബ							
ഭ(2D)																				ഭ	ഭ	ഭ	ഭ							
മ(2E)														മ	മ					മ	മ	മ	മ	മ						
യ(2F)	യ						യ							യ	യ					യ	യ	യ	യ	യ						
ര(30)																				ര			ര							
ല(32)																				ല	ല		ല	ല						
വ(35)																				വ	വ	വ	വ							
ശ(36)						ശ	ശ										ശ			ശ	ശ	ശ	ശ	ശ	ശ					
ഷ(37)	ഷ									ഷ	ഷ			ഷ				ഷ	ഷ	ഷ	ഷ	ഷ	ഷ	ഷ						
സ(38)	സ	സ								സ				സ	സ	സ			സ	സ	സ	സ	സ	സ			സ			സ
ഹ(39)														ഹ						ഹ	ഹ	ഹ	ഹ	ഹ						
ള(33)																												ള		
ഴ(34)	ഴ													ഴ						ഴ			ഴ			ഴ				
റ(31)																				റ	റ		റ							റ