

```
1: #!/usr/bin/env ruby
2: # encoding: utf-8
3:
4: require_relative "phoneme"
5: require "pp"
6:
7: class Numeric
8:   def to_hex; "0x#{self.to_s(16)}" end
9:   def to_bin; "0b#{self.to_s(2)}" end
10: end
11:
12: class Array
13:   def to_hex; "[" + self.map { |e| e.to_hex }.join(", ") + "]" end
14: end
15:
16: class Korean
17:   def initialize()
18:     @first = 0x1100..0x1112
19:     @center = 0x1161..0x1175
20:     @last = 0x11a8..0x11c2
21:     @compat = 0x3130..0x318e
22:     @phoneme = Phoneme.new
23:   end
24:
25:   def transcode(s)
26:     res = ""
27:     i = 0
28:     while i < s.length
29:       ch = 0
30:       if @first.include?(s[i])
31:         use_count = 2
32:         first = s[i] - @first.first
33:         second = s[i + 1] - @center.first
34:         thrid = 0
35:         if i + 2 < s.length and @last.include?(s[i+2])
36:           thrid = s[i+2] - @last.first + 1
37:           use_count += 1
38:         end
39:
40:         ch = (first * 21 + second) * 28 + thrid + 0xac00
41:         i += use_count
42:       else
43:         ch = s[i]
44:         i += 1
45:       end
46:
47:       res += [ch].pack("U")
48:     end
49:
50:     res
51:   end
52:
53:   def transcode_compat(codepoints)
54:     f0 = @phoneme.compact_vowels(codepoints)
55:     f1 = @phoneme.compact_cons(f0)
56:     default_jamo = @phoneme.to_jamo(f1)
57:
58:     return self.transcode(default_jamo)
59:   end
60: end
61:
62: # ok
63: f = File.open("keyboard.dat", "r:utf-8:utf-16le")
64: f.pos = 0x18
65:
66: # ok
67: data = f.read()
68: sentences = data.split "\u0000".encode("utf-16le")
```

```
69:
70: kor = Korean.new
71: sentences.each { |s|
72:   next if s.length < 2
73:   codepoints = s.codepoints.map { |c| c }
74:   pp codepoints.to_hex
75:   pp codepoints.pack("U*").unicode_normalize(:nfd)
76:   pp kor.transcode_compat(codepoints)
77: }
```