

Title: Change proposal on the regular expressions at “Transforming into Standard Korean Syllables”

Date: Feb. 26, 2009

From: Ienup Sung <is@sun.com>

1. Background

In the page 120 of the Unicode Standard Version 5.0, we have:

Transforming into Standard Korean Syllables. A sequence of jamos that do not all match the regular expression for a standard Korean syllable block can be transformed into a sequence of standard Korean syllable blocks by the correct insertion of choseong fillers and jungseong fillers. This transformation of a string of text into standard Korean syllables is performed by determining the syllable breaks as explained in the earlier subsection “Hangul Syllable Boundaries,” then inserting one or two fillers as necessary to transform each syllable into a standard Korean syllable. Thus

$$\begin{aligned} L [^\wedge V] &\rightarrow L V_f [^\wedge V] \\ [^\wedge L] V &\rightarrow [^\wedge L] L_f V \\ [^\wedge V] T &\rightarrow [^\wedge V] L_f V_f T \end{aligned}$$

where $[^\wedge X]$ indicates a character that is not X, or the absence of a character.

There are some possible confusions on the first and the last regular expressions at the above due to that they are not fully self-contained in terms of and by the definition *D119 Standard Korean syllable block* at the same page.

2. Change Proposal

It appears that we can make the regular expressions a bit more self-contained by changing the three lines into the following:

$$\begin{aligned} L [^\wedge (L | V | LV)] &\rightarrow L V_f [^\wedge (L | V | LV)] \\ [^\wedge L] V &\rightarrow [^\wedge L] L_f V \\ [^\wedge (V | LV)] T &\rightarrow [^\wedge (V | LV)] L_f V_f T \end{aligned}$$

If the inclusion of precomposed Hangul syllables at above are not appropriate for the section, then, instead, the following is proposed:

$$\begin{aligned} L [^\wedge (L | V)] &\rightarrow L V_f [^\wedge (L | V)] \\ [^\wedge L] V &\rightarrow [^\wedge L] L_f V \\ [^\wedge V] T &\rightarrow [^\wedge V] L_f V_f T \end{aligned}$$

(Note: Changed are shown with green background.)

END_OF_MEMO.