

To: Unicode Technical Committee
Title: Case pairs across Unicode planes
From: Lorna A. Priest (SIL International)
Date: 26 April 2012

Problem

There are 190 Latin characters in Unicode with a General Category property of “Ll” and no Upper Case Equivalent. 41 of those are “Small Capitals”, and it seems unlikely an upper case would be needed for a Small Capital. An additional 14 have a character decomposition (where the decomposition is to something that already has a case pair). This leaves 135 characters where there is the potential for requiring an upper case character.

Recently, SIL was requested to add a character to our Corporate Private Use Area. This character is an upper case to U+029D LATIN SMALL LETTER J WITH CROSSED-TAIL. This is for a language community in Uganda that has recently finalized their orthography. After making their decision on this character, they learned there was not an upper case in Unicode. We do not anticipate that they will have published materials using this character for some time and thus cannot propose it for addition to Unicode until that time.

This brings us to consider what will happen when the BMP is full. Can (or should) case pairs be assigned across planes (U+029D in the BMP and LATIN CAPITAL LETTER J WITH CROSSED-TAIL in the SMP)? If not, we would like UTC to consider reserving 135 codepoints in the BMP for all potential cases where the upper case could be added to make a pair.