

From: Eric Muller, Adobe Systems
Date: January 30, 2008
Subject: Regression in UAX #14, Line Breaking Properties

It was recently discovered that starting with version 5.0 of UAX #14, there is no longer a line break opportunity between an ideographic period and an alphabetic character, or between an ideographic comma and a digit. Because of the context in which ideographic period and comma appear, it seems desirable to have opportunities at those positions “out of the box”, and indeed previous versions of UAX #14 generated such opportunities. Asmus Freytag submitted an analysis in L2/08-089, and traced this regression to a change introduced in 5.0 to handle situations such as “person(s)”.

The purpose of the present document is not to discuss how to implement the intent of the 5.0 change without causing the regression. Rather, it argues that we need a stronger process to avoid regressions.

Some observations:

- we have to admit that UAX 14 is something we don't know how to fix “on the fly”, that is during the course of a UTC meeting. Even with the best intentions, it is difficult to appreciate all the effects of a change.
- Asmus originally invested a lot of time to ensure that UAX 14 is compatible with Japanese typography, and in fact, multiple members care very much about that. On the other hand, some people care more about Western typography. So far, Asmus managed to satisfy both camps without too many compromises, but it is getting harder.
- we have reached a stage in the life of UAX 14 where a significant number of users depend on UAX 14, such that changes need to be carefully done because regressions are *very* painful.

Some ideas to improve the situation:

- we should develop some method to measure (rather than think about) the impact of a change. I have in mind a test suite with hand-crafted sequences of interest, and/ or a more systematic approach with all possible sequences
- a change should be discussed at one meeting; after that meeting, we can investigate the impact and may be refine a bit the proposal, using the measurements; and approve the change only at the next meeting.
- if the tension between Western and Japanese typography becomes too great, then the Consortium should provide tailorings, so that both worlds can be served. The “Western” vs. “Japanese” is not meant to be exclusive; it should really be “typographies of wide interest”, and for example could eventually include Chinese.

Assuming that the ideas above receive support, I believe that our best course of action is to revert immediately to the pre-5.0 behavior, to implement the process above, and then to reconsider how to deal with “person(s)” using the new process.