To: UTC
From: Mark Davis
Re: Conformance clauses in UAX #29
Date: 2023-04-24

The following is my proposed text for my action 172-A086. I suggest that we incorporate these changes into the proposed update of UAX #29. It has been reviewed by the PAG.

- For a future version of the Unicode Standard, prepare a working draft of a proposed update of UAX #29 with an improved conformance section according to document L2/22-124 item Seg7, taking into account the discussion in UTC #172.

The following changes would be made to https://www.unicode.org/reports/tr29/proposed.html

---

2 Conformance

There are many different ways to divide text elements corresponding to user-perceived characters, words, and sentences, and the Unicode Standard does not restrict the ways in which implementations can produce these divisions. However, it does provide conformance clauses to enable implementations to clearly describe their behavior in relation to the default behavior.

*UAX29-C1. **Extended Grapheme Cluster Boundaries:** An implementation shall choose either **UAX29-C1-1** or **UAX29-C1-2** to determine whether an offset within a sequence of characters is an extended grapheme cluster boundary.*

*UAX29-C1-1. Use property values defined in the Unicode Character Database [UCD] and the rules in **3.1 Default Grapheme Cluster Boundary Specification** to determine the boundaries.*

*UAX29-C1-2. Declare the use of a **profile** of **UAX29-C1-1**, and define that profile with a precise specification of any changes in property values or rules and/or provide a description of programmatic overrides to the behavior of **UAX29-C1-1**.*

*UAX29-C2 **Word Boundaries:** An implementation shall choose either **UAX29-C2-1** or **UAX29-C2-2** to determine whether an offset within a sequence of characters is a word boundary.*

*UAX29-C2-1. Use the property values defined in the Unicode Character Database [UCD] and the rules in **4.1 Default Word Boundary Specification** to determine the boundaries.*

*UAX29-C2-2. Declare the use of a **profile** of **UAX29-C2-1**, and define that profile with a precise specification of any changes in property values or rules and/or provide a description of*

*programmatic overrides to the behavior of **UAX29-C2-1**.*

UAX29-C3 Sentence Boundaries: *An implementation shall choose either **UAX29-C3-1** or **UAX29-C3-2** to determine whether an offset within a sequence of characters is a sentence boundary.*

UAX29-C3-1. *Use the property values defined in the Unicode Character Database [UCD] and the rules in **5.1 Default Sentence Boundary Specification** to determine the boundaries.*

UAX29-C3-2. *Declare the use of a **profile** of **UAX29-C3-1**, and define that profile with a precise specification of any changes in property values or rules and/or provide a description of programmatic overrides to the behavior of **UAX29-C3-1**.*

This specification defines *default* mechanisms; more sophisticated implementations can *and should* tailor them for particular locales or environments and, for the purpose of claiming conformance, document the tailoring in the form of a profile. For example, reliable detection of word boundaries in languages such as Thai, Lao, Chinese, or Japanese requires the use of dictionary lookup or other mechanisms, analogous to English hyphenation. An implementation therefore may need to provide means for a programmatic override of the default mechanisms described in this annex.

Note that a profile can both add and remove boundary positions, compared to the results specified by **UAX29-C1-1, UAX29-C2-1 or UAX29-C3-1.**

> *Notes:*
> - Locale-sensitive boundary specifications, including boundary suppressions, can be expressed in LDML [UTS35]. Profiles are available in the Common Locale Data Repository [CLDR].

…