

**Title:** Proposal to Define a UCD Property for Prefixed Format Control Characters  
**Source:** Laurențiu Iancu (Microsoft Corporation) and Ken Whistler  
**Status:** Individual contribution  
**Action:** For consideration by the Unicode Technical Committee  
**Date:** 2015-10-07

## Abstract

The *prefixed format control characters*, also generically referred to as *subtending marks*, such as U+0600 ARABIC NUMBER SIGN, U+06DD ARABIC END OF AYAH, or U+070F SYRIAC ABBREVIATION MARK, form a class of characters with distinctive characteristics. Wherever there is a need to refer to or manipulate that class of characters, an explicit enumeration is employed, e.g., in the core specification of the Unicode Standard [TUS80] or in the derivation of the Grapheme\_Cluster\_Break property value Prepend proposed for Unicode 9.0 [UAX29]. This document proposes to define a UCD property to handle the characters in that class collectively via properties rather than by hardcoded enumeration.

## Proposal statement

Define a new, informative, binary property named `Prefixed_Control`, with the short name alias `Pre_C`, whose value is Yes (Y) for the following 9 characters as of Unicode Version 8.0:

|         |                          |
|---------|--------------------------|
| U+0600  | ARABIC NUMBER SIGN       |
| U+0601  | ARABIC SIGN SANAH        |
| U+0602  | ARABIC FOOTNOTE MARKER   |
| U+0603  | ARABIC SIGN SAFHA        |
| U+0604  | ARABIC SIGN SAMVAT       |
| U+0605  | ARABIC NUMBER MARK ABOVE |
| U+06DD  | ARABIC END OF AYAH       |
| U+070F  | SYRIAC ABBREVIATION MARK |
| U+110BD | KAITHI NUMBER SIGN       |

## Prefixed format control characters

Prefixed format control characters are proposed marks that are used in combination with a following sequence of digits or letters which need to be formatted as a group to denote a year, a verse number, an abbreviation, or other specialized semantics. Most prefixed controls are Arabic marks, but a few other scripts have adopted similar conventions for certain marks.

Prefixed controls are unusual in that they precede a sequence of characters they interact with instead of following a single base character, as regular combining marks do. For that reason, prefixed controls are not combining marks in the formal sense of that term in the Unicode Standard. Proper display requires specialized rendering support: the glyph of a prefixed control character extends under (subtending), over (supertending), or around (enclosing) the sequence of applicable characters that immediately follow it.

Given their distinctive characteristics, prefixed format control characters form a class whose members can be designated by the proposed `Prefixed_Control` property.

### Characterization of the proposed property

All prefixed format control characters have the `General_Category` property value `Format` (`gc=Cf`) while not being default ignorable (`DI=N`). For additional properties, see the following [Unicode set](#).

The metaproperties of the proposed property are listed below. The suggested property name and short alias are modeled on the existing properties `Bidi_Control` (`Bidi_C`) and `Join_Control` (`Join_C`).

|                     |                               |
|---------------------|-------------------------------|
| Long name           | <code>Prefixed_Control</code> |
| Abbreviated name    | <code>Pre_C</code>            |
| Type                | Binary                        |
| Scope of use        | Display and segmentation      |
| Default value       | No (N)                        |
| Status              | Informative                   |
| Derivational status | Not derivable                 |
| Primary status      | Primary                       |
| Stability           | Not immutable or fixed        |
| Obsolete            | No                            |
| Deprecated          | No                            |
| Stabilized          | No                            |
| Defining data file  | <code>PropList.txt</code>     |
| Documentation       | <a href="#">UAX #44</a>       |
| XML schema          | TBD                           |
| Age                 | 9.0                           |

If the proposed property is accepted, the following specifications and UCD files will need to be updated: the core specification, [UAX #29](#), [UAX #42](#), [UAX #44](#), `PropList.txt`, `PropertyAliases.txt`, `PropertyValueAliases.txt`, and the UCD XML schema and data files.

## References

- [TUS80] The Unicode Consortium, *The Unicode® Standard, Version 8.0.0*, August 2015, <http://www.unicode.org/versions/Unicode8.0.0/>.  
Description of prefixed format control characters in the subsection *Subtending Marks* at <http://www.unicode.org/versions/Unicode8.0.0/ch07.pdf#G27087>.
- [UAX29] Mark Davis, Laurențiu Iancu, *Proposed Update Unicode® Standard Annex #29: Unicode Text Segmentation*, Version 9.0, <http://www.unicode.org/reports/tr29/tr29-28.html>.  
Definition of Grapheme\_Cluster\_Break property value Prepend at <http://www.unicode.org/reports/tr29/tr29-28.html#Prepend>.