

**Subject: Unicode 3.1 Text: Encoding Model for Georgian Script**

**Source: Joe Becker**

**Date: UTC #86 / L2 #183, January 2001 (submitted 26 Jan 2001)**

**Summary:** This document presents the Unicode 3.1 clarification of the encoding model for the Georgian script. The basic encoding is unchanged, and is intended to remain unchanged. However, past errors and omissions in its *description* have led to misinterpretations, which are intended to be remedied by this clarified explanation. This matter of the basic encoding structure is treated separately from other items concerning Georgian, such as additional characters.

---

L2/00-324

Preliminary Minutes of the UTC 85 / L2 182 Joint Meeting  
San Diego, CA -- November 7-10, 2000

**Georgian**

[85-A28] Action Item for Michael Kaplan: Write a document which describes how the current Unicode encoding of Georgian can be used to represent the three Georgian scripts. [L2/00-404]

[85-A29] Action Item for Editorial Committee: Make sure that Unicode V3.1 adequately describes how the current Unicode encoding of Georgian can be used to represent the three Georgian scripts. [L2/00-404]

I have taken over these Actions, many thanks to Michael Kaplan for his draft text and to Michael Everson and Ken Whistler for their contributions ... not implying their endorsement of the final result.

---

## Georgian (revision)

The Georgian script is used primarily for writing the Georgian language and its dialects. It is also used for the Svan and Mingrelian languages, and in the past was used for Abkhaz and other languages of the Caucasus.

**Script Forms.** The Georgian script originates from an inscriptional form called Asomtavruli, from which was derived a manuscript form called Nuskhuri. Together these forms are categorized as Khutsuri (ecclesiastical), but Khutsuri is not itself the name of a script form. Although no longer seen in most modern texts, the Nuskhuri style is still used for liturgical purposes. It was replaced, through a history now uncertain, by an alphabet called Mkhedruli (military), which is now the form used for nearly all modern Georgian writing.

**Case Forms.** The Georgian alphabet is fundamentally caseless, and is used as such in most texts. However, possibly owing to the influence of case forms in other alphabets, modern Georgian is occasionally written with uppercase capital letters. In this typographic departure, it is the Asomtavruli forms that serve to represent uppercase letters, while the lowercase is Mkhedruli or Nuskhuri. This usage parallels the evolution of the Latin alphabet, in which the original linear monumental style came to be considered uppercase, while manuscript styles of the same alphabet came to be represented as lowercase. The Unicode encoding of Georgian follows the Latin analogy: the range U+10A0..U+10CF is used to encode the uppercase capital forms (Asomtavruli), and the basic alphabetic range U+10D0..U+10FF may be regarded as lowercase (Mkhedruli or Nuskhuri). In lowercase (i.e. normal caseless) Georgian text, Mkhedruli or Nuskhuri are distinguished via font, as are regular and italic forms in Latin lowercase.

Font Name	"uppercase" U+10A0..U+10CF	basic/"lower" U+10D0..U+10FF
Mkhedruli	Asomtavruli	Mkhedruli
Nuskhuri	Asomtavruli	Nuskhuri

Because Georgian is predominantly used as a caseless alphabet, no default case mappings are provided for Georgian in the Unicode Character Database. It is inadvisable for generic Unicode text processing to convert Georgian Mkhedruli text to Asomtavruli via a casing operation. In instances where software dealing with Georgian text treats Asomtavruli forms as uppercase letters and requires case folding, this should be done via special casing rules.

**Georgian Paragraph Separator.** The Georgian paragraph separator has a distinct representation, so it has been separately encoded as U+10FB. It visually marks a paragraph end, but it must be followed by a newline character as described in Unicode Standard Annex #13, Unicode Newline Guidelines, to cause a paragraph termination.

**Other Punctuation.** For the Georgian full stop, use U+0589 ARMENIAN FULL STOP.

For additional punctuation to be used with this script, see C0 Controls and ASCII Punctuation (U+0000..U+007F) and General Punctuation (U+2000..U+206F). [END]