

Fraser's Lisu orthography

**WG2/N3326
L2/07-294**

**ISO/IEC JTC1/SC2/WG2
Universal Multiple-Octet Coded Character Set**

Title: Fraser's Lisu orthography: Research notes toward a Unicode encoding

Doc Type: Working Group Document

Author: Richard Cook

Status: Expert Contribution

Action: For consideration by UTC and JTC1/SC2/WG2

Date: 2007-09-15

This document examines questions relating to the encoding of elements of Fraser's (unicameral) Lisu writing system, weighing the benefits of two possible encoding methods, one involving *unification* with (bicameral) Latin Script, and the other involving *disunification*.

After examining the repertory and considering it in relation to encoded (Unicode 5.0, and *Charts Amendment 3* WG2/N3263 [2007-04-27]) characters and case mappings, *unification with Latin Script is proposed as the preferred encoding method*.

Unification of Fraser with Latin Script requires that only 10 characters (out of ~48) be proposed for future encoding. Suitable BMP code points are available in the *Latin Extended-D* block:

A78E LATIN CAPITAL LETTER TURNED B
A78F LATIN CAPITAL LETTER TURNED D
A790 LATIN CAPITAL LETTER TURNED G
A791 LATIN CAPITAL LETTER TURNED J
A792 LATIN CAPITAL LETTER TURNED K
A794 LATIN CAPITAL LETTER TURNED P
A795 LATIN CAPITAL LETTER TURNED R
A796 LATIN CAPITAL LETTER TURNED T
A797 LATIN CAPITAL LETTER TURNED U
A798 LATIN SMALL LETTER TURNED J

Feedback from the technical committees will guide a future encoding proposal.

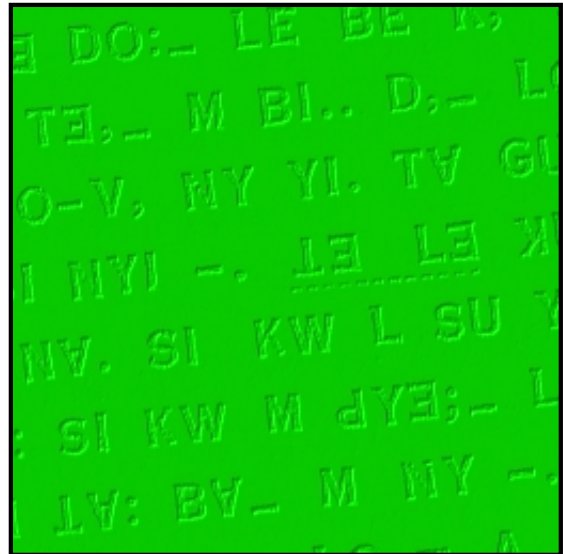
Fraser's Lisu orthography

Research notes toward a Unicode encoding

by

Richard Cook

STEDT, SEI, Unicode



Contents

- 0.0: [Introduction](#)
- 1.0: [Repertory](#)
- 2.0: [Fraser and the UCS](#)
 - 2.1: [Encoded v. Unencoded](#)
 - 2.2: [Case mappings](#)
 - 2.3: [Proposed Encoding Repertory](#)
- 3.0: [Recommendations](#)
- 4.0: [Acknowledgements](#)
- 5.0: [Version History](#)

0.0: Introduction

This document examines questions relating to the encoding of elements of Fraser's (unicameral) Lisu writing system, weighing the benefits of two possible encoding methods, one involving *unification* with (bicameral) Latin Script, and the other involving *disunification*.

After examining the repertory and considering it in relation to encoded (Unicode 5.0, and [Charts Amendment 3 WG2/N3263](#) [2007-04-27]) characters and case mappings, unification with Latin Script is proposed as the preferred encoding method.

1.0: The Fraser Repertory

Fraser letters derive from “LATIN CAPITAL” letters [A..P,R..Z], but Fraser’s Lisu writing is unicameral: there is no case folding in Fraser orthography, and lowercase Latin letters do not occur (except when Fraser is mixed with other orthographies).

In addition to the 25 LATIN CAPITAL letters which occur in their normal form, 15 LATIN CAPITAL letters have forms that are TURNED (i.e. ‘rotated 180 degrees’); 10 other Latin letters which do not contrast (with themselves or with another letter-form in the repertory) when TURNED are used in Fraser writing, but have no TURNED counterpart; only “Q” is completely unused in the writing system. The complete set of 40 Fraser letters is therefore as follows:

```
non-TURNED:  A B C D E F G   J K L       P R T U V           (15)
non-TURNED:                H I       M N O   S       W X Y Z   (10)
TURNED:      A B C D E F G   J K L       P R T U V           (15)
```

A table of the phonological values of these letters when used to write Southern Lisu is given in Bradley’s *Southern Lisu Dictionary* (2005:xxvii); this includes also a list of six tone marks:

```
TONE:   1   2   3   4   5   6
MARK:   .   ,   ..  .,   :   ;
```

In addition, the following four signs, also re-purposing ASCII characters, are also reportedly used:

```
nasal:  '
glide:  _
comma:  -
period: =
```

Treating two of the tone marks as composite, there are a total of (40+4+4=) 48 signs in the Fraser Lisu repertory; if they are not composite, there are (40+6+4=) 50 signs total.

2.0: Fraser and the UCS

2.1: Encoded v. Unencoded

The 25 “non-TURNED” letter forms used in Fraser are derivative of and clearly identical in appearance to the 25 LATIN CAPITAL letters encoded in the Basic Latin block (U+0041..U+0050,U+0052..U+005A).

Of the 15 TURNED letter forms used in Fraser, the following 6 (= 4 [Unicode 5.0] + 2 [Amendment 3]) may

be seen as having encoded equivalents:

0186	LATIN CAPITAL LETTER OPEN O	("LATIN CAPITAL LETTER TURNED C")
018E	LATIN CAPITAL LETTER REVERSED E	("LATIN CAPITAL LETTER TURNED E")
0245	LATIN CAPITAL LETTER TURNED V	("LATIN CAPITAL LETTER TURNED V")
2132	TURNED CAPITAL F	("LATIN CAPITAL LETTER TURNED F")
2C6F	LATIN CAPITAL LETTER TURNED A	(see Charts Amendment 3, p.43)
A780	LATIN CAPITAL LETTER TURNED L	(see Charts Amendment 3, p.53)

The 10 Fraser tone and other marks (simple and composite) are also clearly derivative of and identical in appearance to encoded characters:

0027	APOSTROPHE
002C	COMMA
002D	HYPHEN-MINUS
002E	FULL STOP
003A	COLON
003B	SEMICOLON
003D	EQUALS SIGN

The following table summarizes the encoding status of the 48 Fraser Lisu orthographic elements:

Encoded	
TURNED:	A C E F L V (06)
LATIN:	A B C D E F G J K L P R T U V (15)
LATIN:	H I M N O S W X Y Z (10)
OTHER:	; : , . ' _ - = (08)
Not Encoded	
TURNED:	B D G J K P R T U (09)

2.2: Case mappings

Since Latin Script is bicameral by definition, unifying (nominally or derivationally upper-case) Fraser letters with encoded elements of Latin Script requires that we consider casing relations, even though Fraser itself is unicameral.

Not Encoded	
TURNED:	B D G J K P R T U (09)
CASING:	- - + ? + - + + -

The above 9 TURNED Fraser letters are marked according to whether the corresponding lowercase form is distinctive (“+”) or non-distinctive (“-”) in appearance when TURNED. If the corresponding lowercase form is *distinctive* it is already encoded, in every case but one (marked “?” above); in all other cases it is non-distinctive and unencoded. [Note that an informal rule requiring distinctiveness of TURNED forms is violated in [Charts Amendment 3](#), with encoding of [A781 LATIN SMALL LETTER TURNED L](#).]

Eight case-mapping pairs are as follows (Charts Amendment 3 code points are in **green**; possible *Latin Extended-D* block [U+A720..U+A7FF] code points for new characters are in **red**):

```
0250 LATIN SMALL LETTER TURNED A ↔ 2C6F LATIN CAPITAL LETTER TURNED A
0279 LATIN SMALL LETTER TURNED R ↔ A795 LATIN CAPITAL LETTER TURNED R
0287 LATIN SMALL LETTER TURNED T ↔ A796 LATIN CAPITAL LETTER TURNED T
029E LATIN SMALL LETTER TURNED K ↔ A792 LATIN CAPITAL LETTER TURNED K
1D77 LATIN SMALL LETTER TURNED G ↔ A790 LATIN CAPITAL LETTER TURNED G
A781 LATIN SMALL LETTER TURNED L ↔ A780 LATIN CAPITAL LETTER TURNED L
A798 LATIN SMALL LETTER TURNED J ↔ A791 LATIN CAPITAL LETTER TURNED J
```

By the above analysis, encoding of only one non-Fraser character “LATIN SMALL LETTER TURNED J” is required. These casing relations are also given below in *UnicodeData.txt* format:

```
2C6F;LATIN CAPITAL LETTER TURNED A;Lu;0;L;;;;N;;;0250;
A780;LATIN CAPITAL LETTER TURNED L;Lu;0;L;;;;N;;;A781;
A790;LATIN CAPITAL LETTER TURNED G;Lu;0;L;;;;N;;;1D77;
A791;LATIN CAPITAL LETTER TURNED J;Lu;0;L;;;;N;;;A798;
A792;LATIN CAPITAL LETTER TURNED K;Lu;0;L;;;;N;;;029E;
A795;LATIN CAPITAL LETTER TURNED R;Lu;0;L;;;;N;;;0279;
A796;LATIN CAPITAL LETTER TURNED T;Lu;0;L;;;;N;;;0287;
A798;LATIN SMALL LETTER TURNED J;Ll;0;L;;;;N;;;A791;;
0250;LATIN SMALL LETTER TURNED A;Ll;0;L;;;;N;;;A78D;;
0279;LATIN SMALL LETTER TURNED R;Ll;0;L;;;;N;;;A795;;
0287;LATIN SMALL LETTER TURNED T;Ll;0;L;;;;N;;;A796;;
029E;LATIN SMALL LETTER TURNED K;Ll;0;L;;;;N;;;A792;;
1D77;LATIN SMALL LETTER TURNED G;Ll;0;L;;;;N;;;A790;;
A781;LATIN SMALL LETTER TURNED L;Ll;0;L;;;;N;;;A780;;
```

2.3: Proposed Encoding Repertory

Unification of Fraser with Latin Script requires that only 10 characters (out of the 48 characters in the Fraser repertory) be proposed for future encoding. Suitable BMP code points are available in the *Latin Extended-D* block (U+A720..U+A7FF), in the range **U+A78E..U+A798** (inclusive):

```
A78E LATIN CAPITAL LETTER TURNED B
A78F LATIN CAPITAL LETTER TURNED D
A790 LATIN CAPITAL LETTER TURNED G
A791 LATIN CAPITAL LETTER TURNED J
A792 LATIN CAPITAL LETTER TURNED K
A794 LATIN CAPITAL LETTER TURNED P
A795 LATIN CAPITAL LETTER TURNED R
A796 LATIN CAPITAL LETTER TURNED T
A797 LATIN CAPITAL LETTER TURNED U
A798 LATIN SMALL LETTER TURNED J
```

Note that results of [Charts Amendment 3 WG2/N3263](#) (2007-04-27) were incorporated into this document, removing “LATIN CAPITAL LETTER TURNED A” and “LATIN CAPITAL LETTER TURNED L” from the above list.

2C6F LATIN CAPITAL LETTER TURNED A ([Latin Extended-C](#))
 A780 LATIN CAPITAL LETTER TURNED L ([Latin Extended-D](#))
 A781 LATIN SMALL LETTER TURNED L ([Latin Extended-D](#))

3.0: Recommendations

- It is appropriate to unify Fraser orthographic elements with Latin Script;
 - A total of only (9 “CAPITAL” + 1 “SMALL” =) 10 new characters are required (all proposed in the BMP *Latin Extended-D* block);
 - Of these, 9 new “LATIN CAPITAL LETTER TURNED” characters derive directly from the Fraser set;
 - Case mappings are required in 5 cases (4 of which involve already-encoded “LATIN SMALL” characters);
 - Encoding of non-Fraser “LATIN SMALL LETTER TURNED J” is required for case mapping to “LATIN CAPITAL LETTER TURNED J”;
 - Encoding of 4 other non-Fraser “LATIN SMALL LETTER” characters is *not* required for case mapping, since such “TURNED” forms are not distinctive;
 - Already-encoded characters present no problems in being re-purposed as Fraser orthographic elements (letters, tone marks, etc.);
 - In contrast, disunifying Fraser’s Lisu from Latin Script presents needless duplication, confusion and potential for misuse and abuse.
-

4.0: Acknowledgements

Special thanks to Ken Whistler, for comments and suggestions offered during revisions of this document. Thanks also to Charles Cox, for pointing out the [Charts Amendment 3](#) additions.

5.0: Version History

Last updated: 2007-09-15;11:11:40

Prior Versions: [0](#), [1](#), [2](#).

[Comments?](#) • [TOP](#) • [HOME](#)

