Title:	Comments accompanying the US negative vote on Applications for Registration No. 207 - 225
Source:	NCITS/L2
Date:	August 25, 1998
Action:	Forward to SC2

1. Summary Outline

L2 #	SC2 #	Reg. #	Scope	L2 Position	L2 Comments
98-227	N 3113	207	Irish Gaelic	disapprove	Change to **approve** if images, names are as originally published
98-229	N 3114	208	Ogham	disapprove	Change to **approve** if images, names are as originally published
98-231	N3115	209	Sami suppl. No. 2	disapprove	A-D
98-235	N 3125	212	Ext. Latin	disapprove	F, 2.1
98-236	N 3126	213	Suppl. Minor European	disapprove	A-D, H, 2.2
98-237	N 3127	214	Ext. Cyrillic	disapprove	F
98-238	N 3128	215	Greek	disapprove	F, 2.3
98-239	N 3129	216	Ext. African	disapprove	F, 2.4
98-240	N 3130	217	Math. pt. 1	disapprove	A-D, 2.5
98-241	N 3131	218	Math. pt. 2	disapprove	A-D, 2.6
98-242	N 3132	219	Hebrew. pt. 1	disapprove	A-D, 2.7
98-243	N 3133	220	Hebrew. pt. 2	disapprove	A-D, 2.8
98-244	N 3134	221	Armenian	disapprove	A-D, 2.9
98-245	N 3135	222	Georgian	disapprove	A-D, 2.10
98-246	N 3136	223	Ext. non-Slavic Cyrillic	disapprove	A-D, 2.11
98-247	N 3137	224	Ext. Arabic	disapprove	A-D, H
98-248	N 3138	225	Ext. Latin	disapprove	A-D, 2.12
98-253	N 3140	210	Sami No. 1	disapprove	A, E, G, 12.13
98-254	N 3141	211	Sami No. 2	disapprove	A, E, G, 12.14

General Comments

A. The US is generally opposed to further registrations for 7 and 8 bit character sets.

B. When SC2 approves registration of a character set, what is registered should include an image of the code table as it was originally published, not a version with redrawn glyphs.

C. When SC2 approves registration of a character set, the names of characters should be taken from the character set as it was originally published.

D. Renaming of characters with ISO/IEC 10646 names implies mapping of characters in the set being registered to ISO/IEC 10646 characters. Such mapping has not been reviewed by NCITS/L2 and should not be sanctioned via character set registration. Just because a character in a character set being registered has the same name as an ISO/IEC character, it should not be assumed that the two characters are identical.

E. The registration includes mapping to ISO/IEC 10646 characters, both explicitly and by use of ISO/IEC 10646 names. Such mapping should not be sanctioned via character set registration; it needs separate review by qualified experts

Comments on Registration Status

F. This character set is already registered. See Section 3 for specifics.

G. Registration is not needed because the character set is "not intended to be used to be used in conjunction with any other graphic character set, through code extension techniques according to ISO/IEC 2022 or ISO/IEC 4873, or otherwise."

Comment on Ballot Status:

H. Mappings are given for characters currently under ballot without identifying their status.

2.   Comments on Individual Applications for Registration

A blank cell in the "Comments" column indicates a typographical error.

2.1   L2/98-235, SC2 N3125 (Registration No 212, Extension of the Latin alphabet coded character set for bibliographic interchange):

Incorrect mappings:

Code	Name	Maps to	Not to	Comments
02/07	Section sign or paragraph mark	U+00A7	U+00B6	Image in code chart is §, so mapping should be to U+00A7
02/08	Prime	U+2032	U+2033	
02/15	Registered trade mark	U+00AE	U+2122	For consistency with character 02/10 in document N 3138
03/00	Ayn	U+02BF	U+02BD	
04/12	High inverted comma centered	U+0313	U+0312	For consistency with character 07/14 in document N 3138

Document SC2 N3138 is an application for registration for ANSI/NISO Z39.47:1993 *Extended Latin alphabet coded character set for bibliographic use.* Many characters in ISO 5426:1980, *Extension of the Latin alphabet coded character set for bibliographic interchange* and ANSI/NISO Z39.47:1993 have a common origin in the USMARC Extended Latin Character Set, published by the Library of Congress in 1968. Characters common to ISO 5426:1980 and ANSI/NISO Z39.47:1993 should have identical mappings.

04/08 Trema, Diaeresis and 04/09 Umlaut

ISO 5426 contains two characters, 04/08 Trema, Diaeresis and 04/09 Umlaut, which correspond to a single character, U+0308 COMBINING DIAERESIS, in ISO/IEC 10646.

There are three possible options for this many:one situation:

a) Unify 04/08 and 04/09 by mapping both to U+0308 (the choice for the European library CHASE project);
b) Map one of the pair to U+0308, and the other to a Private Use Area value;
c) Map 04/08 to U+0308, and 04/09 to the proposed character COMBINING LATIN SMALL LETTER E ABOVE, i.e., the antique form of the *umlaut*.

2.2   L2/98-236, SC2 N3126 (Registration No 213, Supplementary minor European and obsolete typographical Latin set):

Incorrect mappings:

Code	Name	Maps to	Not to	Comments
02/02	SIX SPOKED ASTERISK	U+2736		Per CHASE mapping
02/07	LATIN SMALL LETTER SLOPED D	U+03B4	U+2202	Variant of small delta
03/02	REFERENCE MARK	U+203B	U+20B3	
04/07	COMBINING LATIN SMALL LETTER O ABOVE	Not U+030A	U+030A	U+030A is COMBINING RING ABOVE, i.e., not the letter o
06/01	LATIN CAPITAL LETTER G WITH CEDILLA	U+0122	U+01E4	Code chart image shows cedilla
06/09	LATIN CAPITAL LETTER YR	U+01A6		U+01A6, LATIN LETTER YR, is an uppercase letter
07/09	LATIN SMALL LETTER YR	Not U+01A6	U+01A6	See preceding comment

2.3   L2/98-238, SC2 N3128 (Registration No 215, Greek alphabet coded character set for bibliographic information interchange):

Incorrect mappings:

Code	Name	Maps to	Not to	Comments
04/01	Alpha (capital letter)	U+0391	U+0390	
04/02	Beta (capital letter)	U+0392	U+0391	
04/01	Alpha (small letter)	U+03B1	U+03B0	
04/02	Beta (small letter)	U+03B2	U+03B1	

2.4 L2/98-239, SC2 N3129 (Registration No 216, Extended African Latin alphabet coded character set for bibliographic information interchange):

The identification of the character set proposed for registration is incorrect. The date in the ISO number should be *1983* (not *1996*). The correct title is *Documentation -- African coded character set for bibliographic information interchange*, not *Extended African Latin alphabet coded character set for bibliographic information interchange*.

Incorrect mappings:

Code	Name	Maps to	Not to	Comments
03/08	Mid-central vowel; schwa (small)	U+01DD	U+0259	See below

LATIN SMALL LETTER REVERSED E, the new name given to character 03/08 in the Application for registration is in error. U+0259 is LATIN SMALL LETTER SCHWA. However, the correct mapping is to U+01DD LATIN SMALL LETTER TURNED E, which case pairs with U+018E, the mapping for 02/08.

Incorrect name:

03/15 (Open high front vowel, small) is named LATIN SMALL LETTER I WITH STROKE but mapped to U+026A, LATIN LETTER SMALL CAPITAL I. The mapping is correct.

2.5 L2/98-240, SC2 N3130 (Registration No 217, Mathematical coded character set for bibliographic information interchange, part 1):

Incorrect mappings:

Code	Name	Maps to	Not to	Comments
02/05	Circle, overlay	U+20DD	U+20DB	
03/02	Equivalent to	U+223C	U+2236	
03/13	Left angle bracket	U+2329	U+3008	See below
04/13	Angle bracket, right	U+232A	U+3009	See below
05/11	Left arrow over right arrow	U+21C6	U+22C6	
05/12	Functional relationship	U+21A6	U+22A6	

Most of the characters in part 1 of ISO 6862:1996 *Information and documentation -- Mathematical coded character set for bibliographic information interchange* are derived from the *Maths* character set developed by the British Library. The British Library is a participant in the CHASE project to establish Unicode/UCS mappings for the repertoires of character sets used by European libraries. The CHASE mappings, to U+2329 and U+232A, should be used.

2.6 L2/98-241, SC2 N3131 (Registration No 218, Mathematical coded character set for bibliographic information interchange, part 2):

Incorrect mappings:

Code	Name	Maps to	Not to	Comments
03/08	Long triangle	????	U+22B2	See below
04/08	Long triangle, underlined	????	U+22B4	See below
04/13	Equiangular	U+225A	U+22AF	
04/14	Implies	U+25B9	U+22B9	
04/15	Hamilton operator	U+25BF	U+22BF	
06/00	Magnitude of	U+007C	U+0076	
06/15	Spherical angle	U+2222	U+2221	

In ISO 6862:1996, *Information and documentation -- Mathematical coded character set for bibliographic information interchange,* the characters 03/08 and 04/08 are named *Long triangle* and *Long triangle, underlined* respectively, with no indication of mathematical functionality. The mappings in SC2 N3131 assign mathematical functionality to these characters (*normal subgroup of* and *normal subgroup of or equal to* respectively).

Unifications

US opinion is that the following characters should be treated as duplicates, and mapped as follows:

Code	Name	Map to	Cf. In part 1
05/03	Vector or sum	U+2228	07/03, Logical or
05/04	Sum or union of classes or sets	U+222A	05/04, Union of sets between limits
05/05	Is included in set	U+2282	05/01, Proper inclusion in set
06/03	Includes in set	U+2227	07/04, Logical and
06/04	Product of intersection of classes or sets	U+2229	06/04, Intersection of classes or sets between limits
06/05	Includes in set	U+2283	06/01, Properly includes in set

2.7 L2/98-242, SC2 N3132 (Registration No 219, Hebrew alphabet coded character set for bibliographic information interchange, part 1):

Incorrect mapping:

Code	Name	Maps to	Not to	Comments
03/10	HEBREW PUNCTUATION SOF PASUQ	U+003A	U+05C3	See below

The name of character 03/10 is incorrect. Part 1 of ISO 8957:1996, *Information and documentation -- Hebrew alphabet coded character set for bibliographic information interchange* is derived from the USMARC Hebrew character set, where the corresponding character is COLON. The *sof pasuq* is encoded in Part 2 of ISO 8957:1996, at position 02/02.

2.8 L2/98-243, SC2 N3133 (Registration No 220, Hebrew alphabet coded character set for bibliographic information interchange, part 2):

Incorrect mapping:

Code	Name	Maps to	Not to	Comments
04/11	HEBREW ACCENT TEVIR	????	U+059B	Names same, but images different
05/11	HEBREW ACCENT RAFE	????	U+05BF	See below

This mapping unifies the HEBREW ACCENT RAFE (5B in Part 2) and the HEBREW POINT RAFE (4C in Part 1). These characters have different names and images, and should not be unified.

The US questions the mapping of ancient Hebrew cantillation marks to combining diacritical marks with no supporting evidence.

Code	Name	Code	Name
41	HEBREW ACCENT TSERE	0308	COMBINING DIAERESIS
47	HEBREW ACCENT DOUBLE ACUTE	030B	COMBINING DOUBLE ACUTE ACCENT
50	HEBREW ACCENT QAMATS	030D	COMBINING VERTICAL LINE ABOVE
51	HEBREW ACCENT ACUTE	0301	COMBINING ACUTE ACCENT
52	HEBREW ACCENT GRAVE	0300	COMBINING GRAVE ACCENT
55	HEBREW ACCENT SAMARIAN HOLAM	0302	COMBINING CIRCUMFLEX ACCENT
56	HEBREW ACCENT SAMARIAN SEGOL	030C	COMBINING CARON

2.9 L2/98-244, SC2 N3134 (Registration No 221, Armenian alphabet coded character set for bibliographic information interchange):

Incorrect mappings:

Code	Name	Maps to	Not to	Comments
04/15	ARMENIAN ABBREVIATION MARK	U+055F	U+055C	
07/10	QUOTATION MARK	U+2033	U+0022	Mapping specified in SC2/WG2 N1616
07/12	ARMENIAN SEMICOLON	U+00B7	U+0387	Mapping specified in SC2/WG2 N1616

2.10 L2/98-245, SC2 N3135 (Registration No 222, Georgian alphabet coded character set for bibliographic information interchange):

Incorrect mapping:

Code	Name	Maps to	Not to	Comments
04/14	GEORGIAN COMMA	U+00B7	U+0387	

2.11   L2/98-246, SC2 N3136 (Registration No 223, Extended non-Slavic Cyrillic alphabet coded character set for bibliographic information interchange):

The proposal shows ISO 10586 (Georgian character set), not ISO 10754.

2.12   L2/98-248, SC2 N3138 (Registration No 225, Extension of the Latin alphabet coded character set for bibliographic interchange):

- The title of this application for registration is that same as application for registration No, 212, but the applications are for different character sets.
- The US recommends as the title for this application *Extended Latin alphabet coded character set for bibliographic use,* corresponding to the title of ANSI/NISO Z39.47-1993.
- The mapping value for 07/08 (*right cedilla* in ANSI/NISO Z39.47-1993) should be U+031C, in accordance with the mapping for the equivalent character in the USMARC Extended Latin set.

2.13   L2/98-253, SC2 N3140 (Registration No 210, Sami complete 8-bit graphic character set no. 1

The application states that the character set is "not intended to be used to be used in conjunction with any other graphic character set, through code extension techniques according to ISO/IEC 2022 or ISO/IEC 4873, or otherwise." Given this, NCITS/L2 cannot understand why an application for registration was even initiated (causing work for the SC2 Secretariat and national standards bodies).

This character set is intended for use "primarily in Windows applications." The application for registration presents no evidence that Microsoft Corporation has either endorsed or used the character set. The application for registration does not reference a Windows code page designation.

No evidence is presented in the application that this character set existed prior to the application itself. If this is the case, and the application constitutes the initial specification for this character set, NCITS/L2 considers this to be inappropriate use of the registration process. NCITS/L2 does not wish to see the registration process used as a way to "standardize" a character set that does not conform to ISO/IEC provisions.

NCITS/L2 is gravely concerned that there is a danger that the character set registration process could be used to create new and uncoordinated Windows code pages by providing a mechanism for their identification and publication. Such a prospect is the antithesis of standardization, and NCITS/L2 cannot believe that SC2 would sanction this.

NCITS/L2 welcomes assurances from the National Body of Sweden that our fears are groundless.

2.14   L2/98-254, SC2 N3141 (Registration No 211, Sami complete 8-bit graphic character set no. 2

The application states that the character set is "not intended to be used to be used in conjunction with any other graphic character set, through code extension techniques according to ISO/IEC 2022 or ISO/IEC 4873, or otherwise." Given this, NCITS/L2 cannot understand why an application for registration was even initiated (causing work for the SC2 Secretariat and national standards bodies).

Macintosh code pages are proprietary encodings published by Apple Computer, Inc. The application for registration presents no evidence that Apple Computer, Inc. has been consulted about the character set, nor whether the company has consented to the inclusion of its trademarked logo.

To avoid the trademark issue, NCITS/L2 recommends that position 15/00 in this character set be designated "This position shall not be used" instead of "APPLE LOGOTYPE."

Apple Computer does not register its own Macintosh encodings, so escape sequences for the most widely used Macintosh encodings have not been assigned by ISO. It is therefore pointless to approve Application for Registration No 211.

No evidence is presented in the application that this character set existed prior to the application itself. If this is the case, and the application constitutes the initial specification for this character set, NCITS/L2 considers this to be inappropriate use of the registration process. NCITS/L2 does not wish to see the registration process used as a way to "standardize" a character set that does not conform to ISO/IEC provisions and would be proprietary if published by Apple Computer.

NCITS/L2 is gravely concerned that there is a danger that the character set registration process could be used to create new and uncoordinated Macintosh encodings by providing a mechanism for their identification and publication. Such a prospect is the antithesis of standardization, and NCITS/L2 cannot believe that SC2 would sanction this.

NCITS/L2 welcomes assurances from the National Body of Sweden that our fears are groundless.

3.   Appendix: Registrations in Effect

The final characters of the respective escape sequences which are already assigned are shown in this table:

No.	Character Set	Scope	Final Character (F)	Graphically
212	ISO 5426:1980	Extended Latin	5/0	P
214	ISO 5427:1984	Extended Cyrillic	5/1	Q
215	ISO 5428:1980	Greek	5/3	S
216	ISO 6438:1983	African	4/13	M

This information would have been available to SC2 if the standards as published had been submitted for registration instead of redrawn versions. The escape sequences are specified in Clause 3.2 of each standard.

4.   Additional Comment

When these registrations were reviewed, NCITS/L2 was unaware that the International Federation of Library Associations (IFLA) had issued Update 2 to the *UNIMARC Manual: Bibliographic Format*. In this Update, codes are assigned to four additional character sets, thus authorizing their use as G0 through G3 sets in UNIMARC records.

Transcription of the section from *Update 2: Errata and Minor Changes*

100 $a/26-29 Character Sets. Add

07 = 1SO 10586 (Georgian set)
08 = ISO 8957 (Hebrew set)
09 = ISO 5426-2 (Latin characters used in minor European languages and obsolete typography)
50 = ISO 10646 Level 3 (Unicode)

Note that ISO 10646, being a 16-bit character set, contains all necessary characters. When positions 26-27 contains '50' this will be used for C0, C1 and all g sets. Positions 28-33 will contain blanks.

When a UNIMARC record contains more than four different graphic character sets, UNIMARC identifies the additional graphic sets via escape sequences (identical to those assigned by ISO registration procedures). Update 2 does not specify escape sequences for any of the newly added sets.

The action of the Permanent UNIMARC Committee to allow more character sets to be used in UNIMARC records reinforces Comments B, C, and D. A registration must unequivocally identify the character set *as published*, not a modified or abbreviated version, and mappings to UCS characters (if part of the registration) must be properly reviewed.

The specification of "ISO 10646 Level 3 (Unicode)" as a UNIMARC character set is an indication that the international library community intends to use ISO/IEC 10646 encoding in UNIMARC records instead of 7- and 8-bit character sets with ISO 2022 techniques eventually.