

Reference number of working document: **ISO/IEC JTC1/SC22/WG20 N634**

Date: 1998-12-21

Reference number of document: **ISO/IEC FCD2 14652**

Committee identification: ISO/IEC JTC1/SC22

Secretariat: ANSI

## **Information technology —**

### **Specification method for cultural conventions**

*Technologies de l'information —*

*Méthode de modélisation des conventions culturelles*

1

**Document type:** International standard

**Document subtype:** if applicable

**Document stage:** (40) Enquiry

**Document language:** E

		Page
2	Contents	
3		
4	1 SCOPE	1
5	2 NORMATIVE REFERENCES	1
6	3 TERMS, DEFINITIONS AND NOTATIONS	2
7	4 FDCC-set	6
8	4.0 FDCC-set definition	6
9	4.1 LC_IDENTIFICATION	10
10	4.2 LC_CTYPE	11
11	4.3 LC_COLLATE	27
12	4.4 LC_MONETARY	42
13	4.5 LC_NUMERIC	46
14	4.6 LC_TIME	47
15	4.7 LC_MESSAGES	53
16	4.8 LC_PAPER	53
17	4.9 LC_NAME	55
18	4.10 LC_ADDRESS	57
19	4.11 LC_TELEPHONE	57
20	5 CHARMAP	58
21	6 REPERTOIREMAP	62
22	7 CONFORMANCE	90
23	Annex A (informative) DIFFERENCES FROM POSIX	91
24	Annex B (informative) RATIONALE	93
25	Annex C (informative) BNF GRAMMAR	107
26	Annex D (informative) INDEX	112
27	BIBLIOGRAPHY	115
28		

29 **FOREWORD**

30

31 ISO (the International Organization for Standardization) and IEC (the International  
32 Electrotechnical Commission) form the specialized system for worldwide standardization.  
33 National bodies that are members of ISO or IEC participate in the development of  
34 International Standards through technical committees established by the respective  
35 organization to deal with particular fields of technical activity. ISO and IEC technical  
36 committees collaborate in fields of mutual interest. Other international organizations,  
37 governmental and non-governmental, in liaison with ISO and IEC, also take part in the  
38 work.

39

40 International Standards are drafted in accordance with the rules given in the ISO/IEC  
41 Directives, Part 3.

42

43 In the field of information technology, ISO and IEC have established a joint technical  
44 committee, ISO/IEC JTC 1. Draft International Standards adopted by the joint technical  
45 committee are circulated to national bodies for voting. Publication as an International  
46 Standard requires approval by at least 75 % of the national bodies casting a vote.

47

48 International Standard ISO/IEC 14652 was prepared by Joint Technical Committee  
49 ISO/IEC JTC 1., "Information Technology", subcommittee 22, "Programming languages,  
50 their environments and system software interfaces".

51

52 The Standard extends the concept of the locale specifications defined primarily in  
53 subclauses 2.4 and 2.5 of ISO/IEC 9945-2:1993 "Information Technology - Portable  
54 Operating System Interface (POSIX) - Part 2: Shell and Utilities". The major extensions  
55 from this text are listed in annex A.

56

57 The annexes A, B, C and D are for information only.

58

## 59 Introduction

60

61 This International Standard defines a general mechanism to specify cultural conventions,  
62 and it defines formats for a number of specific cultural conventions in the areas of  
63 character classification and conversion, sorting, number formatting, monetary formatting,  
64 date formatting, message display, paper formats, addressing of persons, postal address  
65 formatting, telephone number handling, measurement handling, and a way to specify how  
66 much is covered and the status of it.

67

68 There are a number of benefits coming from this standard:

69

70 Rigid specification

Using this International Standard, a user can rigidly  
specify a number of the cultural conventions that apply  
to the information technology environment of the user.

71

72

73

74 Cultural adaptability

If an application has been designed and built in a  
cultural neutral manner, the application may use the  
specifications as data to its APIs, and thus the same  
application may accommodate different users in a  
culturally acceptable way to each of the users, without  
change of the binary application.

75

76

77

78

79

80

81 Internationalization

An internationalized application needs to be designed  
and implemented as cultural neutral, so that, at run time,  
it draws on the cultural conventions of the user thus  
giving the application the ability to support cultural  
conventions of many different cultures. This standard  
specifies those cultural conventions and how to specify  
data for them. With internationalized applications the  
application developer is relieved from getting the  
different information to support all the cultural  
environments for the expected customers of the product.  
The application developer is thus ensured of culturally  
correct behaviour as specified by the customer, and  
possibly more markets may be reached as customers may  
have the possibility to provide the data themselves for  
markets that were not targeted.

82

83

84

85

86

87

88

89

90

91

92

93

94

95

96

97

98 Uniform behaviour

When an application has been internationalized, it is  
dependent on the operating system support for  
internationalization what level of service is available to  
the user. Some operating systems do not provide  
internationalization support. Some provide a set of  
cultural conventions, that the user then can chose the  
most suitable from. Yet others provide the capability that  
users may be allowed to supply their own cultural  
convention specifications. Using the cultural  
conventions thus available, users may enjoy consistent  
and correct behaviour on these issues from the  
internationalized applications.

99

100

101

102

103

104

105

106

107

108

109 The specification format is independent of platforms and specific encoding, and targeted to  
110 be usable from a wide range of programming languages. It is expected that the primary  
111 areas of use is within the POSIX operating system, although it could also be used in other  
112 environments.

113

114 A number of cultural conventions, such as spelling, hyphenation rules and terminology,  
115 and classification of characters such as Japanese gaiji characters, are not specifiable with  
116 this standard, but the standard provides mechanisms to define new categories and also new  
117 keywords within existing categories. An internationalized application may take advantage  
118 of information provided with the FDCC-set (such as the language) to provide further  
119 internationalized services to the user.

120

121 This International Standard defines a format compatible with the one used in the  
122 International String Ordering standard, ISO/IEC 14651. This International Standard is  
123 backwards compatible with the ISO/IEC 9945:1993 POSIX shell and utilities standard, and  
124 it has enhanced functionality in a number of areas such as ISO/IEC 10646 support, more  
125 classification of characters, transliteration, dual currency support, enhanced date and time  
126 formatting, paper size identification, personal name writing, postal address formatting,  
127 telephone number handling, and management of categories. There is enhanced support for  
128 character sets including ISO 2022 handling and an enhanced method to separate the  
129 specification of cultural conventions from an actual encoding via a description of the  
130 character repertoire employed. A standard set of values for all the categories has been  
131 defined covering the repertoire of ISO/IEC 10646.

# Information technology — Specification method for cultural conventions

## 1 SCOPE

This International Standard specifies a description format for the specification of cultural conventions, a description format for character sets, and a description format for binding character names to ISO/IEC 10646, plus a set of default values for some of these items.

The specification is upward compatible with POSIX locale specifications - a locale conformant to POSIX specifications will also be conformant to the specifications in this Standard, while the reverse condition will not hold. The descriptions is intended to also be of use in other systems than POSIX. The descriptions are intended to be coded in text files to be used via Application Programming Interfaces, that are expected to be developed for a number of programming languages.

## 2 NORMATIVE REFERENCES

The following normative documents contain provisions which, through reference in this text, constitute provisions of this International Standard. For dated references, subsequent amendments to, or revisions of, any of these publications do not apply. However, parties to agreements based on this International Standard are encouraged to investigate the possibility of applying the most recent editions of the normative documents indicated below. For undated references, the latest edition of the normative document referred to applies. Members of ISO and IEC maintain registers of currently valid International Standards.

ISO 639 (all parts), *Code for the representation of names of languages*.

ISO/IEC 2022, *Information technology - Character code structure and extension techniques*.

ISO 3166 (all parts), *Code for the representation of names of countries*.

ISO 4217 (all parts), *Codes for the representation of currencies and funds*.

ISO 8601, *Data elements and interchange formats - Information interchange - Representation of dates and times*.

ISO/IEC 9945-2:1993, *Information technology - Portable Operating System Interface (POSIX) Part 2: Shell and Utilities*.

ISO/IEC 10646:1997, *Information technology - Universal Multiple-Octet Coded Character Set (UCS), including Cor.1 and AMD 1-9*.

ISO/IEC 14651, *Information technology - International string ordering - Method for comparing character strings and description of a default tailorable ordering*.

ISO/IEC 15897:1998, *Procedures for registration of cultural conventions*.

181  
182  
183  
184  
185  
186  
187  
188  
189  
190  
191  
192  
193  
194  
195  
196  
197  
198  
199  
200  
201  
202  
203  
204  
205  
206  
207  
208  
209  
210  
211  
212  
213  
214  
215  
216  
217  
218  
219  
220  
221  
222  
223  
224  
225  
226  
227  
228  
229  
230

### 3 TERMS, DEFINITIONS AND NOTATIONS

#### 3.1 Terms and definitions

For the purposes of this International Standard, the terms and definitions given in the following apply.

**3.1.1 byte:** An individually addressable unit of data storage that is equal to or larger than an octet, used to store a character or a portion of a character.

A byte is composed of a contiguous sequence of bits, the number of which is application defined. The least significant bit is called the low-order bit; the most significant bit is called the high-order bit.

**3.1.2 character:** A member of a set of elements used for the organization, control or representation of data.

**3.1.3 coded character:** A sequence of one or more bytes representing a single character.

**3.1.4 text file:** A file that contains characters organized into one or more lines.

**3.1.5 cultural convention:** A data item for information technology that may vary dependent on language, territory, or other cultural habits.

**3.1.6 FDCC-set:** A Set of Formal Definitions of Cultural Conventions. The definition of the subset of a user's information technology environment that depends on language and cultural conventions. Note: the FDCC-set is a superset of the "locale" term in C and POSIX.

**3.1.7 charmap:** A definition of a mapping between symbolic character names and character codes, plus related information"

**3.1.8 repertoiremap:** A definition of a mapping between symbolic character names and characters for the repertoire of characters used in a FDCC-set, further described in clause 6.

**3.1.9 character class:** A named set of characters sharing an attribute associated with the name of the class.

**3.1.10 collation:** The logical ordering of strings according to defined precedence rules.

**3.1.11 collating element:** The smallest entity used to determine logical ordering.

See collating sequence. A collating element shall consist of either a single character, or two or more characters collating as a single entity. The LC\_COLLATE category in the associated FDCC-set determines the set of collating elements.

**3.1.12 multicharacter collating element:** A sequence of two or more characters that collate as an entity.

231 For example, in some languages two characters are sorted as one letter, as in the case for  
 232 Danish and Norwegian "aa".

233

234 **3.1.13 collating sequence:** The relative order of collating elements as determined by the  
 235 setting of the LC\_COLLATE category in the applied FDCC-set.

236

237 **3.1.14 equivalence class:** A set of collating elements with the same primary collation  
 238 weight.

239

240 Elements in an equivalence class are typically elements that naturally group together, such  
 241 as all accented letters based on the same letter.

242

243 The collation order of elements within an equivalence class is determined by the weights  
 244 assigned on any subsequent levels after the primary weight.

245

246 **3.1.15 affirmative response:** A string conforming to the definition of LC\_MESSAGES  
 247 category keyword "yesexpr", giving the string values that is acceptable as an affirmative  
 248 response to a question.

249

250 **3.1.16 negative response:** A string conforming to the definition of LC\_MESSAGES  
 251 category keyword "noexpr", giving the string values that is acceptable as a negative  
 252 response to a question.

253

## 254 3.2 Notations

255

256 The following notations and common conventions for specifications apply to this standard:

257

### 258 3.2.1 Notation for defining syntax

259

260 In this standard, the description of an individual record in a FDCC-set is done using the  
 261 syntax notation given in the following.

262

263 The syntax notation looks as follows:

264

265 "<format>",[<arg1>,<arg2>,....,<argn>]

266

267 The <format> is given in a format string enclosed in double quotes, followed by a number  
 268 of parameters, separated by commas. It is similar to the format specification defined in  
 269 clause 2.12 in the POSIX-2 standard and the format specification used in C language  
 270 printf() function. The format of each parameter is given by an escape sequence as follows:

271

272	%s	specifies a string
273	%d	specifies a decimal integer
274	%c	specifies a character
275	%o	specifies an octal integer
276	%x	specifies a hexadecimal integer

277

278 A " " (an empty character position) in the syntax string represent one or more <blank>  
 279 characters.

280



281 All other characters in the format string except

282

283       %%       specifies a single %

284       \n       specifies an end-of-line

285

286 represent themselves.

287

288 The notation "... " is used to specify that repetition of the previous specification is optional,  
289 and this is done in both the format string and in the parameter list.

290

291

### 292 3.2.2 Continuation of lines

293

294 A line in a specification can be continued by placing an escape character as the last visible  
295 graphic character on the line; this continuation character shall be discarded from the input.  
296 Comment lines shall not be continued on a subsequent line using an escaped <newline>.

297

### 298 3.2.3 Portable character set

299

300 The following table defines the characters in the portable character set and the  
301 corresponding symbolic character names used to identify each character in a character  
302 description text.

303

304

305 Table 1: portable character set

306

307 Symbolic name	308 Glyph	309 UCS	310 UCS name
311 <NUL>		<U0000>	NULL (NUL)
312 <alert>		<U0007>	BELL (BEL)
313 <backspace>		<U0008>	BACKSPACE (BS)
314 <tab>		<U0009>	CHARACTER TABULATION (HT)
315 <carriage-return>		<U000D>	CARRIAGE RETURN (CR)
316 <newline>		<U000A>	LINE FEED (LF)
317 <vertical-tab>		<U000B>	LINE TABULATION (VT)
318 <form-feed>		<U000C>	FORM FEED (FF)
319 <space>		<U0020>	SPACE
320 <exclamation-mark>	!	<U0021>	EXCLAMATION MARK
321 <quotation-mark>	"	<U0022>	QUOTATION MARK
322 <number-sign>	#	<U0023>	NUMBER SIGN
323 <dollar-sign>	\$	<U0024>	DOLLAR SIGN
324 <percent-sign>	%	<U0025>	PERCENT SIGN
325 <ampersand>	&	<U0026>	AMPERSAND
326 <apostrophe>	'	<U0027>	APOSTROPHE
327 <left-parenthesis>	(	<U0028>	LEFT PARENTHESIS
328 <right-parenthesis>	)	<U0029>	RIGHT PARENTHESIS
329 <asterisk>	*	<U002A>	ASTERISK
330 <plus-sign>	+	<U002B>	PLUS SIGN
331 <comma>	,	<U002C>	COMMA
332 <hyphen-minus>	-	<U002D>	HYPHEN-MINUS
333 <hyphen>	-	<U002D>	HYPHEN-MINUS
334 <full-stop>	.	<U002E>	FULL STOP
335 <period>	.	<U002E>	FULL STOP
336 <slash>	/	<U002F>	SOLIDUS
337 <solidus>	/	<U002F>	SOLIDUS
338 <zero>	0	<U0030>	DIGIT ZERO
339 <one>	1	<U0031>	DIGIT ONE
340 <two>	2	<U0032>	DIGIT TWO
341 <three>	3	<U0033>	DIGIT THREE
342 <four>	4	<U0034>	DIGIT FOUR
343 <five>	5	<U0035>	DIGIT FIVE
344 <six>	6	<U0036>	DIGIT SIX
345 <seven>	7	<U0037>	DIGIT SEVEN
	8	<U0038>	DIGIT EIGHT
	9	<U0039>	DIGIT NINE

346	<colon>	:	<U003A>	COLON
347	<semicolon>	;	<U003B>	SEMICOLON
348	<less-than-sign>	<	<U003C>	LESS-THAN SIGN
349	<equals-sign>	=	<U003D>	EQUALS SIGN
350	<greater-than-sign>	>	<U003E>	GREATER-THAN SIGN
351	<question-mark>	?	<U003F>	QUESTION MARK
352	<commercial-at>	@	<U0040>	COMMERCIAL AT
353	<A>	A	<U0041>	LATIN CAPITAL LETTER A
354	<B>	B	<U0042>	LATIN CAPITAL LETTER B
355	<C>	C	<U0043>	LATIN CAPITAL LETTER C
356	<D>	D	<U0044>	LATIN CAPITAL LETTER D
357	<E>	E	<U0045>	LATIN CAPITAL LETTER E
358	<F>	F	<U0046>	LATIN CAPITAL LETTER F
359	<G>	G	<U0047>	LATIN CAPITAL LETTER G
360	<H>	H	<U0048>	LATIN CAPITAL LETTER H
361	<I>	I	<U0049>	LATIN CAPITAL LETTER I
362	<J>	J	<U004A>	LATIN CAPITAL LETTER J
363	<K>	K	<U004B>	LATIN CAPITAL LETTER K
364	<L>	L	<U004C>	LATIN CAPITAL LETTER L
365	<M>	M	<U004D>	LATIN CAPITAL LETTER M
366	<N>	N	<U004E>	LATIN CAPITAL LETTER N
367	<O>	O	<U004F>	LATIN CAPITAL LETTER O
368	<P>	P	<U0050>	LATIN CAPITAL LETTER P
369	<Q>	Q	<U0051>	LATIN CAPITAL LETTER Q
370	<R>	R	<U0052>	LATIN CAPITAL LETTER R
371	<S>	S	<U0053>	LATIN CAPITAL LETTER S
372	<T>	T	<U0054>	LATIN CAPITAL LETTER T
373	<U>	U	<U0055>	LATIN CAPITAL LETTER U
374	<V>	V	<U0056>	LATIN CAPITAL LETTER V
375	<W>	W	<U0057>	LATIN CAPITAL LETTER W
376	<X>	X	<U0058>	LATIN CAPITAL LETTER X
377	<Y>	Y	<U0059>	LATIN CAPITAL LETTER Y
378	<Z>	Z	<U005A>	LATIN CAPITAL LETTER Z
379	<left-square-bracket>	[	<U005B>	LEFT SQUARE BRACKET
380	<backslash>	\	<U005C>	REVERSE SOLIDUS
381	<reverse-solidus>	\	<U005C>	REVERSE SOLIDUS
382	<right-square-bracket>	]	<U005D>	RIGHT SQUARE BRACKET
383	<circumflex-accent>	^	<U005E>	CIRCUMFLEX ACCENT
384	<circumflex>	^	<U005E>	CIRCUMFLEX ACCENT
385	<low-line>	_	<U005F>	LOW LINE
386	<underscore>	_	<U005F>	LOW LINE
387	<grave-accent>	`	<U0060>	GRAVE ACCENT
388	<a>	a	<U0061>	LATIN SMALL LETTER A
389	<b>	b	<U0062>	LATIN SMALL LETTER B
390	<c>	c	<U0063>	LATIN SMALL LETTER C
391	<d>	d	<U0064>	LATIN SMALL LETTER D
392	<e>	e	<U0065>	LATIN SMALL LETTER E
393	<f>	f	<U0066>	LATIN SMALL LETTER F
394	<g>	g	<U0067>	LATIN SMALL LETTER G
395	<h>	h	<U0068>	LATIN SMALL LETTER H
396	<i>	i	<U0069>	LATIN SMALL LETTER I
397	<j>	j	<U006A>	LATIN SMALL LETTER J
398	<k>	k	<U006B>	LATIN SMALL LETTER K
399	<l>	l	<U006C>	LATIN SMALL LETTER L
400	<m>	m	<U006D>	LATIN SMALL LETTER M
401	<n>	n	<U006E>	LATIN SMALL LETTER N
402	<o>	o	<U006F>	LATIN SMALL LETTER O
403	<p>	p	<U0070>	LATIN SMALL LETTER P
404	<q>	q	<U0071>	LATIN SMALL LETTER Q
405	<r>	r	<U0072>	LATIN SMALL LETTER R
406	<s>	s	<U0073>	LATIN SMALL LETTER S
407	<t>	t	<U0074>	LATIN SMALL LETTER T
408	<u>	u	<U0075>	LATIN SMALL LETTER U
409	<v>	v	<U0076>	LATIN SMALL LETTER V
410	<w>	w	<U0077>	LATIN SMALL LETTER W
411	<x>	x	<U0078>	LATIN SMALL LETTER X
412	<y>	y	<U0079>	LATIN SMALL LETTER Y
413	<z>	z	<U007A>	LATIN SMALL LETTER Z
414	<left-brace>	{	<U007B>	LEFT CURLY BRACKET
415	<left-curly-bracket>	{	<U007B>	LEFT CURLY BRACKET
416	<vertical-line>		<U007C>	VERTICAL LINE
417	<right-brace>	}	<U007D>	RIGHT CURLY BRACKET
418	<right-curly-bracket>	}	<U007D>	RIGHT CURLY BRACKET
419	<tilde>	~	<U007E>	TILDE

420

421 This standard places only the following requirements on the encoded values of the

422 characters in the portable character set:

423

424 (1) The encoded values associated with each member of the portable character set  
425 shall be invariant across all FDCC-sets supported by the application. If this is not the case,  
426 the results achieved by an application accessing those FDCC-sets are unspecified.

427

428 (2) The encoded values associated with the digits '0' to '9' shall be such that the  
429 value of each character after '0' shall be one greater than the value of the previous  
430 character.

431

432 The standard may use other symbolic character names than the above in examples, to  
433 illustrate the use of the range of symbols allowed by the syntax specified in 4.0.1.

434

#### 435 **4 FDCC-set**

436

437 A FDCC-set is the definition of the subset of a user's information technology environment  
438 that depends on language and cultural conventions. It is made up from one or more  
439 categories. Each category is identified by its name and controls specific aspects of the  
440 behaviour of components of the system. This standard defines the following categories:

441

442	LC_IDENTIFICATION	Versions and status of categories
443	LC_CTYPE	Character classification, case conversion and code 444 transformation.
445	LC_COLLATE	Collation order.
446	LC_TIME	Date and time formats.
447	LC_NUMERIC	Numeric, non-monetary formatting.
448	LC_MONETARY	Monetary formatting.
449	LC_MESSAGES	Formats of informative and diagnostic messages and 450 interactive responses.
451	LC_PAPER	Paper format
452	LC_NAME	Format of writing personal names
453	LC_ADDRESS	Format of postal addresses
454	LC_TELEPHONE	Format for telephone numbers, and other telephone 455 information

456

457 In future editions of this standards further categories may be added. Other category names  
458 beginning with the 3 characters "LC\_" are intended for future standardization, except for  
459 category names beginning with the five characters "LC\_X\_" which shall not be used for  
460 future addition of categories specified in this International Standard. An application may  
461 thus use category names beginning with the five characters "LC\_X\_" for application  
462 defined categories to avoid clashes with future standardized categories.

463

464 This standard also defines an FDCC-set named "i18n" with values for each of the above  
465 categories.

466

#### 467 **4.0 FDCC-set definition**

468

469 FDCC-sets are described with the syntax presented in this subclause. For the purposes of  
470 this standard, the text is referred to as the FDCC-set definition text or FDCC-set source  
471 text.

472 The **FDCC-set definition text** shall contain one or more FDCC-set category source  
473 definitions, and shall not contain more than one definition for the same FDCC-set  
474 category. If the text contains source definitions for more than one category, application-  
475 defined categories, if present, shall appear after the categories defined by this clause. A  
476 category source definition shall contain either the definition of a category or a copy  
477 directive. In the event that some of the information for a FDCC-set category, as specified  
478 in this standard, is missing from the FDCC-set source definition, the behaviour of that  
479 category, if it is referenced, is unspecified. A FDCC-set category is the normal way of  
480 specifying a single FDCC.

481  
482 There are no **naming conventions** for FDCC-sets specified in this international standard,  
483 but ISO/IEC 15897:1998 specifies naming rules for POSIX locales, charmaps and  
484 repertoiremaps, that may also be applied to FDCC-sets, charmaps and repertoiremaps  
485 specified according to this standard.

486  
487 A **category source definition** shall consist of a category header, a category body, and a  
488 category trailer. A category header shall consist of the character string naming of the  
489 category, beginning with the characters "LC\_". The category trailer shall consist of the  
490 string "END", followed by one or more "blank"s and the string used in the corresponding  
491 category header.

492  
493 The **category body** shall consist of one or more lines of text. Each line shall contain an  
494 identifier, optionally followed by one or more operands. Identifiers shall be either  
495 keywords, identifying a particular FDCC, or collating elements, or section symbols, or  
496 transliteration statements. In addition to the keywords defined in this standard, the source  
497 can contain application-defined keywords. Each **keyword** within a category shall have a  
498 unique name (i.e., two categories can have a commonly-named keyword); no keyword  
499 shall start with the characters "LC\_". Identifiers shall be separated from the operands by  
500 one or more "blank"s.

501  
502 **Operands** shall be characters, collating elements, section symbols, or strings of characters.  
503 Strings shall be enclosed in double-quotes. Literal double-quotes within strings shall be  
504 preceded by the <escape character>, described below. When a keyword is followed by  
505 more than one operand, the operands shall be separated by semicolons; "blank"s shall be  
506 allowed before and/or after a semicolon.

507  
508

#### 509 **4.0.1 Character representation**

510

511 Individual characters, characters in strings, and collating elements shall be represented  
512 using symbolic names, UCS notation or characters themselves, or as octal, hexadecimal, or  
513 decimal constants as defined below. When constant notation is used, the resultant  
514 FDCC-set definitions need not be portable between systems.

515

516 (0) The left angle bracket (<) is a reserved symbol, denoting the  
517 start of a symbolic name; when used to represent itself it  
518 shall be preceded by the escape character.

519

520 (1) A character can be represented via a **symbolic name**,  
521 enclosed within angle brackets (< and >). The symbolic

522 name, including the angle brackets, shall exactly match a  
 523 symbolic name defined in a charmap or a repertoiremap to  
 524 be used, and shall be replaced by a character value  
 525 determined from the value associated with the symbolic  
 526 name in the charmap or a value associated via a  
 527 repertoiremap. Repertoiremaps have predefined symbolic  
 528 names for UCS characters, see clause 6. A FDCC-set may  
 529 also use the UCS notation of clause 6 to represent characters,  
 530 without a repertoiremap being defined for the FDCC-set. Use  
 531 of the escape character or a right angle bracket within a  
 532 symbolic name shall be invalid unless the character is  
 533 preceded by the escape character.

534 Example: <c>;<c-cedilla> "<M><a><y>"

535

536

537 The items (2), (3), (4) and (5) are deprecated and are retained for compatibility with the  
 538 POSIX standard. FDCC-sets should be specified in a coded character set independent way,  
 539 using symbolic names. To make actual use of the FDCC-set, it shall be used together with  
 540 charmaps and/or repertoiremaps, so that the symbolic character names can be resolved into  
 541 the actual character encoding used.

542

543

(2) A character can be represented by the character itself, in  
 544 which case the value of the character is application-defined.  
 545 Within a string, the double-quote character, the escape  
 546 character, and the right angle bracket character shall be  
 547 escaped (preceded by the escape character) to be interpreted  
 548 as the character itself. Outside strings, the characters

549

550

, ; < > escape\_char

551

552 shall be escaped to be interpreted as the character itself.

553

554

Example: c ä "May"

555

556

(3) A character can be represented as an octal constant. An octal  
 557 constant shall be specified as the escape character followed  
 558 by two or more octal digits. Each constant shall represent a  
 559 byte value.

560

561

Example: \143; \347; "\115"

562

563

(4) A character can be represented as a hexadecimal constant. A  
 564 hexadecimal constant shall be specified as the escape  
 565 character followed by an x followed by two or more  
 566 hexadecimal digits. Each constant shall represent a byte  
 567 value.

568

569

Example: \x63;\xe7;

570

571

(5) A character can be represented as a decimal constant. A  
 572 decimal constant shall be specified as the escape character

573 followed by a d followed by two or more decimal digits.  
 574 Each constant shall represent a byte value.

575  
 576 Example: \d99; \d231;

577  
 578 (6) Multibyte characters can be represented by concatenated  
 579 constants specified in byte order with the last constant  
 580 specifying the least significant byte of the character.  
 581 Concatenated constants can include a mix of the above  
 582 character representations.

583  
 584 Example: \143\xe7; "\115\xe7\d171"

585  
 586 Only characters existing in the character set for which the FDCC-set definition is created  
 587 shall be specified, whether using symbolic names, the characters themselves, or octal,  
 588 decimal, or hexadecimal constants. If a charmap is present, only characters defined in the  
 589 charmap can be specified using octal, decimal, or hexadecimal constants. Symbolic names  
 590 not present in the charmap can be specified and shall be ignored, as specified under item  
 591 (1) above.

## 592 **4.0.2 Pre-category statements**

593  
 594  
 595 In a FDCC-set the following statements can precede category specifications, and they  
 596 apply to all categories in the specified FDCC-set.

### 597 **4.0.2.1 comment\_char**

598  
 599  
 600 The following line in a FDCC-set modifies the comment character. It shall have the  
 601 following syntax, starting in column 1:

602  
 603 "comment\_char %c\n", <comment character>

604  
 605 The comment character shall default to the number-sign (#). All examples in this standard  
 606 use "%" as the <comment char>, except where otherwise noted. Blank lines and lines  
 607 containing the <comment char> in the first position, and the remainder of a line with a  
 608 <comment char> occurring where an end of line may occur, shall be ignored.

### 609 **4.0.2.2 escape\_char**

610  
 611  
 612 The following line in a FDCC-set modifies the escape character to be used in the text. It  
 613 shall have the following syntax, starting in column 1:

614  
 615 "escape\_char %c\n", <escape character>

616  
 617 The escape character shall default to backslash "\". All examples in this standard uses "/"  
 618 as the escape character, except where otherwise noted.

### 619 **4.0.2.3 repertoiremap**

620  
 621  
 622 The following line in a FDCC-set specifies the name of a repertoiremap used to define the

623 symbolic character names in the FDCC-set. There may be at most one "repertoiremap"  
624 line. It shall have the following syntax, starting in column 1:

625  
626 "repertoiremap %s\n", <repertoiremap>

#### 627 628 **4.0.2.4 charmap**

629  
630 The following line in a FDCC-set specifies the name of a charmap which may be used  
631 with the FDCC-set. It shall have the following syntax, starting in column 1:

632  
633 "charmap %s\n", <charmap>

634  
635 There may be more than one charmap specification in a FDCC-set. For the actual use of a  
636 FDCC-set, at most one charmap may be in use for a given instantiation. FDCC-sets are  
637 recommended to be character encoding independent, and the "charmap" keyword gives a  
638 hint on which charmaps that a FDCC-set is meant to be supported by, so that for example  
639 all day and month names can be represented using the charmaps specified. As FDCC-sets  
640 may be character encoding independent, other charmaps than the ones specified may be  
641 used with the FDCC-set.

### 642 643 644 **4.1 LC\_IDENTIFICATION**

645  
646 The LC\_IDENTIFICATION category defines properties of the FDCC-set, and which  
647 specification methods the FDCC-set is conforming to. All keywords are mandatory unless  
648 otherwise noted, and the operands are strings. The following keywords shall be defined:

649		
650	title	Title of the FDCC-set.
651	source	Organization name of provider of the source.
652	address	Organization postal address.
653	contact	Name of contact person.
654	email	Electronic mail address of the organization, or contact
655		person.
656	tel	Telephone number for the organization, in international
657		format.
658	fax	Fax number for the organization, in international format.
659	language	Natural language to which the FDCC-set applies, as specified
660		in ISO 639.
661	territory	The geographic extent where the FDCC-set applies (need not
662		be a national extent), as two-letter form of ISO 3166.
663	audience	If not for general use, an indication of the intended user
664		audience. This keyword is optional.
665	application	If for use of a special application, a description of the
666		application. This keyword is optional.
667	abbreviation	Short name for provider of the source. This keyword is
668		optional.
669	revision	Revision number consisting of digits and zero or more full
670		stops (".").
671	date	Revision date in the format according to this example:
672		"1995-02-05" meaning the 5th of February, 1995.

673 If any of the above information is non-existent, it must be stated in each case; the  
 674 corresponding string is then the empty string. If required information is not present in ISO  
 675 639 or ISO 3166, the relevant Maintenance Authority should be approached to get the  
 676 needed item registered.

677

678 Note: Only one culture can be addressed with the concepts of a FDCC-set; to address  
 679 for example a bilingual culture, one need to have 2 FDCC-sets.

680

681 **category** Shall be used to define that a category is present and what  
 682 specification the category is claiming conformance to. The  
 683 first operand is a string in double-quotes that describes the  
 684 specification that the category is claiming conformance to,  
 685 and the following values shall be defined:  
 686 "i18n:1999"  
 687 "posix:1993"  
 688 The second operand is a string with the category name,  
 689 where the category names of clause 4 shall be defined. More  
 690 than one "category" keyword may be given, but only one per  
 691 category name.

692

693 The "i18n" LC\_IDENTIFICATION category is:

694

```

695 LC_IDENTIFICATION
696 % This is the ISO/IEC 14652 "i18n" definition for
697 % the LC_IDENTIFICATION category.
698 %
699 title "ISO/IEC 14652 i18n FDCC-set"
700 source "ISO/IEC JTC1/SC22/WG20 - internationalization"
701 address "C/o Keld Simonsen, Skt. Jorgens Alle 8, DK-1615
702 Kobenhavn V"
703 contact "Keld Simonsen"
704 email "keld@dkuug.dk"
705 tel "+45 3122-6543"
706 fax "+45 3325-6543"
707 language ""
708 territory "ISO"
709 revision "1.0"
710 date "1997-12-20"
711 %
712 category "i18n:1999";LC_IDENTIFICATION
713 category "i18n:1999";LC_CTYPE
714 category "i18n:1999";LC_COLLATE
715 category "i18n:1999";LC_TIME
716 category "i18n:1999";LC_NUMERIC
717 category "i18n:1999";LC_MONETARY
718 category "i18n:1999";LC_MESSAGES
719 category "i18n:1999";LC_PAPER
720 category "i18n:1999";LC_NAME
721 category "i18n:1999";LC_ADDRESS
722 category "i18n:1999";LC_TELEPHONE
723
724 END LC_IDENTIFICATION
725

```

726

727

## 4.2 LC\_CTYPE

728

729 The LC\_CTYPE category defines character classification, case conversion, character  
 730 transformation, and other character attribute mappings. Support for the portable character  
 731 set is required.

732

733 A series of characters in a specification can be represented by the hexadecimal symbolic



734 ellipsis symbol "." (two dots), the decimal symbolic ellipses symbols "..." (4 dots), the  
735 double increment hexadecimal symbolic ellipses "..(2).", or the absolute ellipses "... (3  
736 dots).

737

738 The **hexadecimal symbolic ellipsis** ("..") specification is only valid between symbolic  
739 character names. The symbolic names shall consist of zero or more nonnumeric characters  
740 from the set shown with visible glyphs in Table 1, followed by an integer formed by one  
741 or more hexadecimal digits, using uppercase letters only for the range "A" to "F". The  
742 characters preceding the hexadecimal integer shall be identical in the two symbolic names,  
743 and the integer formed by the hexadecimal digits in the second symbolic name shall be  
744 identical to or greater than the integer formed by the hexadecimal digits in the first name.  
745 This shall be interpreted as a series of symbolic names formed from the common part and  
746 each of the integers in hexadecimal format using uppercase letters only between the first  
747 and the second integer, inclusive, and with a length of the symbolic names generated that  
748 is equal to the length of the first (and also the second) symbolic name. As an example,  
749 <U010E>..750 and <U0111>, in that order.

751

752 The **decimal symbolic ellipsis** ("...") specification is only valid between symbolic  
753 character names. The symbolic names shall consist of zero or more nonnumeric characters  
754 from the set shown with visible glyphs in Table 1, followed by an integer formed by one  
755 or more decimal digits. The characters preceding the decimal integer shall be identical in  
756 the two symbolic names, and the integer formed by the decimal digits in the second  
757 symbolic name shall be identical to or greater than the integer formed by the decimal  
758 digits in the first name. This shall be interpreted as a series of symbolic names formed  
759 from the common part and each of the integers in decimal format between the first and the  
760 second integer, inclusive, and with a length of the symbolic names generated that is equal  
761 to the length of the first (and also the second) symbolic name. As an example,  
762 <j0101>...<j0104> is interpreted as the symbolic names <j0101>, <j0102>, <j0103>, and  
763 <j0104>, in that order.

764

765 The **double increment hexadecimal symbolic ellipses** ("..(2).") works like the  
766 hexadecimal symbolic ellipses, but generates only every other of the symbolic character  
767 names. As an example. <U01AC>..(2)..<U01B2> is interpreted as the symbolic character  
768 names <U01AC>, <U01AE>, <U01B0>, and <U01B2>, in that order.

769

770 The **absolute ellipsis** specification is only valid within a single encoded character set. An  
771 ellipsis shall be interpreted as including in the list all characters with an encoded value  
772 higher than the encoded value of the character preceding the ellipsis and lower than the  
773 encoded value of the character following the ellipsis. The absolute ellipsis specification is  
774 deprecated, as this is only relevant to FDCC-sets not using symbolic characters.  
775 As an example, \x30;...;\x39 includes in the character class all characters with encoded  
776 values between the endpoints.

777

#### 778 4.2.1 Basic keywords

779

780 The following keywords shall be defined. In the descriptions, the term "automatically  
781 included" means that it shall not be an error to either include the referenced characters or  
782 to omit them; the interpreting system shall provide them if missing and accept them  
783 silently if present.

784	<b>copy</b>	Specify the name of an existing FDCC-set to be used as the source for the
785		definition of this category. If this keyword is specified, no other keyword
786		shall be specified.
787	<b>upper</b>	Define characters to be classified as uppercase letters. No character
788		specified for the keywords "cntrl", "digit", "punct", or "space" shall be
789		specified. The uppercase letters A through Z of the portable character set,
790		shall automatically belong to this class, with application-defined character
791		values. The keyword may be omitted.
792	<b>lower</b>	Define characters to be classified as lowercase letters. No character
793		specified for the keywords "cntrl", "digit", "punct", or "space" shall be
794		specified. The lowercase letters a through z of the portable character set,
795		shall automatically belong to this class, with application-defined character
796		values. The keyword may be omitted.
797	<b>alpha</b>	Define characters to be classified as used to generate word-like identifiers
798		for natural languages; such as letters, syllabic or ideographic characters. No
799		character specified for the keywords "cntrl", "digit", "punct", or "space"
800		shall be specified. In addition, characters classified as either "upper" or
801		"lower" shall automatically belong to this class. The keyword may be
802		omitted.
803	<b>digit</b>	Define the characters to be classified as numeric digits. Digits
804		corresponding to the values 0, 1, 2, 3, 4, 5, 6, 7, 8, and 9 can be specified
805		in groups of 10 digits, and in ascending order of the values they represent.
806		The digits of the portable character set are automatically included. If this
807		keyword is not specified, the digits 0 through 9 of the portable character set
808		shall automatically belong to this class, with application-defined character
809		values. The "digit" keyword is used to specify which characters are
810		accepted as digits in input, and should list digits used with all the scripts
811		supported by the FDCC-set. The keyword may be omitted.
812	<b>outdigit</b>	Define the characters to be classified as numeric digits for output. Digits
813		corresponding to the values <0>, <1>, <2>, <3>, <4>, <5>, <6>, <7>, <8>,
814		and <9> can be specified, and in ascending order of the values they
815		represent. The intended use is for all places where digits are used for
816		output, including numeric and monetary formatting, and date and time
817		formatting. Only one set of 10 digits may be specified. If this keyword is
818		not specified, the digits 0 through 9 of the portable character set shall
819		automatically belong to this class, with application-defined character values.
820		The keyword may be omitted.
821	<b>blank</b>	Define characters to be classified as "blank" characters. If this keyword is
822		unspecified, the characters <space> and <tab>, with application-defined
823		character values, shall belong to this character class.
824	<b>space</b>	Define characters to be classified as white-space characters, to find
825		syntactical boundaries. No character specified for the keywords "upper",
826		"lower", "alpha", "digit", "graph", or "xdigit" shall be specified. If this
827		keyword is not specified, the characters <space>, <form-feed>, <newline>,
828		<carriage-return>, <tab>, and <vertical-tab>, shall automatically belong to
829		this class, with application-defined character values. Any characters
830		included in the class "blank" shall be automatically included. The class
831		should not include the NO-BREAK spaces characters <U00A0>, <U2007>,
832		<UFEFF>, as these characters should not be used for word boundaries. The
833		keyword may be omitted.

834	<b>cntrl</b>	Define characters to be classified as control characters. No character
835		specified for the keywords "upper", "lower", "alpha", "digit", "punct",
836		"graph", "print", or "xdigit" shall be specified. The keyword shall be
837		specified.
838	<b>punct</b>	Define characters to be classified as punctuation characters. No character
839		specified for the keywords "upper", "lower", "alpha", "digit", "cntrl",
840		"xdigit", or as the <space> character shall be specified. The keyword shall
841		be specified.
842	<b>xdigit</b>	Define the characters to be classified as hexadecimal digits. Only the
843		characters defined for the class "digit" shall be specified, in ascending
844		sequence by numerical value, followed by one or more sets of six characters
845		representing the hexadecimal digits 10 through 15, with each set in
846		ascending order (for example <A>, <B>, <C>, <D>, <E>, <F>, <a>, <b>,
847		<c>, <d>, <e>, <f>). If this keyword is not specified, the digits <0> through
848		<9>, the uppercase letters "A" through <F>, and the lowercase letters <a>
849		through <f>, shall automatically belong to this class, with application-
850		defined character values.
851	<b>graph</b>	Define characters to be classified as printable characters, not including the
852		<space> character. If this keyword is not specified, characters specified for
853		the keywords "upper", "lower", "alpha", "digit", "xdigit", and "punct" shall
854		belong to this character class. No character specified for the keyword "cntrl"
855		shall be specified.
856	<b>print</b>	Define characters to be classified as printable characters, including the
857		<space> character. If this keyword is not provided, characters specified for
858		the keywords upper, lower, alpha, digit, xdigit, punct, graph, and the
859		<space> character shall belong to this character class. No character
860		specified for the keyword "cntrl" shall be specified.
861	<b>toupper</b>	Define the mapping of lowercase letters to uppercase letters. The operand
862		shall consist of character pairs, separated by semicolons. The characters in
863		each character pair shall be separated by a comma and the pair enclosed by
864		parentheses. The first character in each pair shall be the lowercase letter, the
865		second the corresponding uppercase letter. Only characters specified for the
866		keywords "lower" and "upper" shall be specified. If this keyword is not
867		specified, the lowercase letters <a> through <z>, and their corresponding
868		uppercase letters <A> through <Z>, shall automatically be included, with
869		application-defined character values.
870	<b>tolower</b>	Define the mapping of uppercase letters to lowercase letters. The operand
871		shall consist of character pairs, separated by semicolons. The characters in
872		each character pair are separated by a comma and the pair enclosed by
873		parentheses. The first character in each pair shall be the uppercase letter, the
874		second the corresponding lowercase letter. Only characters specified for the
875		keywords "lower" and "upper" shall be specified. If this keyword is speci-
876		fied, the uppercase letters <A> through <Z>, and their corresponding
877		lowercase letter, shall be specified. If this keyword is not specified, the
878		mapping shall be the reverse mapping of the one specified for toupper.
879	<b>class</b>	Define characters to be classified in the class with the name given in the
880		first operand, which is a string. This string shall only contain characters of
881		the portable character set that either has the string "LETTER" in its
882		description, digits and <hyphen-minus> and <low-line> that all appear in
883		the portable character set. The following operands are characters. This

884 keyword is optional. The keyword can only be specified once per named  
 885 class.

886 **combining** Characters to form composite graphic symbols, such  
 887 as characters listed in ISO/IEC 10646:1993 annex B.1.

888 **combining\_level3** Characters to form composite graphic symbols, that  
 889 may also be represented by other characters, such as  
 890 characters listed in ISO/IEC 10646-1:1993 annex B.2.

891 The class names "upper", "lower", "alpha", "digit", "space", "cntrl", "punct",  
 892 "graph", "print", "xdigit", and "blank" are taken to mean the classes defined  
 893 by the respective keywords.

894 **map** Define the mapping of characters. The first operand is a string, defining the  
 895 name of the mapping. The string shall only contain letters, digits and  
 896 <hyphen-minus> and <low-line> from the portable character set. The  
 897 following operands shall consist of character pairs, separated by semicolons.  
 898 The characters in each character pair shall be separated by a comma and the  
 899 pair enclosed by parentheses. The first character in each pair shall be the  
 900 character to map from, the second the corresponding character to map to.  
 901 This keyword is optional. The keyword can only be specified once per  
 902 named mapping.

903

904 The mapping names "toupper", and "tolower" are taken to mean the  
 905 mapping defined by the respective keywords.

906

907 Example of use of the "map" keyword:

908

```
909 map "kana",(<U30AB>,<U304B>);<U30AC>,<U304C>);<U30AD>,<U304D>)
```

910

911 This example introduces a new mapping "kana" that maps three Katakana characters to corresponding Hiragana  
 912 characters.

913

914 Table 2 shows the allowed character class combinations.

915

916

917 Table 2: Valid Character Class Combinations

918

919 Class	upper	lower	alpha	digit	space	cntrl	punct	graph	print	xdigit	blank
921 upper		+	A	x	x	x	x	A	A	+	x
922 lower	+		A	x	x	x	x	A	A	+	x
923 alpha	+	+		x	x	x	x	A	A	+	x
924 digit	x	x	x		x	x	x	A	A	A	x
925 space	x	x	x	x		+	*	*	*	x	+
926 cntrl	x	x	x	x	+		x	x	x	x	+
927 punct	x	x	x	x	+	x		A	A	x	+
928 graph	+	+	+	+	+	x	+		A	+	+
929 print	+	+	+	+	+	x	+	+		+	+
930 xdigit	+	+	+	+	x	x	x	A	A		x
931 blank	x	x	x	x	A	+	*	*	*	x	

932

933 NOTES:

934 Note 1: Explanation of codes:

- 935 A Automatically included; see text  
 936 + Permitted  
 937 x Mutually exclusive  
 938 \* See note 2

939  
 940 Note 2: The <space> character, which is part of the "space" and "blank" class, cannot  
 941 belong to "punct" or "graph", but automatically shall belong to the "print" class. Other  
 942 "space" or "blank" characters can be classified as "punct", "graph", and/or "print".

## 943 944 **4.2.2 Character string transliteration**

945  
 946 The following keywords may be used to transliterate strings, by transforming substrings in  
 947 the source to substrings in the target string. The capabilities are limited to simple  
 948 transliteration based on substring substitution, while more advanced transliteration  
 949 schemes, for example based on pattern matching, is either cumbersome to specify, or not  
 950 addressed. The transliteration may for example be from the Cyrillic script to the Latin  
 951 script. Transliteration is often language dependent, and the language to be transliterated to  
 952 is identified with the FDCC-set, which may also be used to identify a specific language to  
 953 be transliterated from. Transliteration may also be to a specific repertoire of characters,  
 954 determined for example by limitations of displaying equipment, or what the user can  
 955 intelligibly read. The capabilities here allows for multiple fallback, so that the specification  
 956 can be valid for all target character repertoires, eliminating the need for specific data for  
 957 each target repertoire. Transliteration of an incoming character string to a character string  
 958 in a FDCC-set can be specified with the following keywords and transliteration statements.

959 960 961 <b>translit_start</b>	The "translit_start" keyword is followed by one or more transliteration statements assigning character transliteration values to transliterating elements, and include statements copying transliteration specifications from other FDCC-sets.
962 963 964 <b>translit_end</b>	The end of the transliteration statements.
965 966 <b>include</b>	The name of the FDCC-set in text form to transliterate from, and the repertoiremap for the FDCC-set to be used for the definition of the transliteration statements. Other transliteration statements may follow to replace specification of the copied FDCC-set. This keyword is optional.
967 968 969 970 <b>default_missing</b>	defines one or more characters to be used if no transliteration statement can be applied to a input <transliteration-source>.

### 971 972 973 **4.2.2.1 Transliteration statements**

974  
 975 The "translit\_start" keyword may be followed by transliteration statements. The syntax for  
 976 a transliteration statement is:

977  
 978  
 979 `"%s %s;%s;...;%s\n",<transliteration-source>,<transliteration-string>,...`

980  
 981 Each <transliteration-source> shall consist of one or more characters (in any of the forms  
 982 defined in 4.0.1). The <transliteration-source> that is the longest in terms of number of  
 983 characters that match the input string is the one selected for transliteration.

984

985 If a transliteration statement contains more than one <transliteration-string>, the order that  
 986 each <transliteration-string> occurs in the transliteration statement defines the precedence  
 987 order for choosing a particular <transliteration-string> to substitute for the <transliteration-  
 988 source>. When a process makes use of a transliteration statement to transliterate text, and  
 989 that transliteration statement contains more than one <transliteration-string>, that process  
 990 shall choose the first <transliteration-string>, in the defined precedence order, that satisfies  
 991 the requirements of the transliteration.

992  
 993 Note: the exact definition of the concept of satisfying the requirements of the  
 994 transliteration is outside the context of this standard. If, for example, a  
 995 transliteration involves a change in the coded character set of a string, a  
 996 <transliteration-string> must be chosen, all of whose elements are members of that  
 997 coded character set. In order to determine this, it would be expected that a  
 998 repertoire describing which characters are to be present in the resulting transformed  
 999 string be available to the transliteration API. Also, a transliteration may involve  
 1000 requirements such as that string length not change under transliteration. Such  
 1001 requirements may also affect the choice among alternative <transliteration-string>  
 1002 values.

1003  
 1004 If more than one transliteration statement is given for a given <transliteration-source> this  
 1005 is an error, and duplicate transliteration statements are ignored. Tailoring of transliteration  
 1006 statements may be done via the "redefine" keyword.

#### 1007 1008 **4.2.2.2 "include" keyword**

1009  
 1010 The "include" keyword specifies a set of transliteration statements in text form to be  
 1011 included in the applied transliteration.

1012  
 1013 The syntax of the "include" statement is:

```
1014  
1015 "include %s;%s\n", <FDCC-set>, <repertoiremap>
```

1016  
 1017 <FDCC-set> is a string identifying the FDCC-set to be included from.

1018  
 1019 <repertoiremap> is a string identifying the repertoiremap used in the FDCC-set being  
 1020 included, and is used to map character specifications from the specified FDCC-set into the  
 1021 current FDCC-set.

#### 1022 1023 **4.2.2.3 Example of use of transliteration**

```
1024  
1025 translit_start  
1026 include "de_DE"; "de_repmap"  
1027 default_missing "<?>"  
1028 <ae> <a:>; <e*>; "<a><e>"; "<e>"  
1029 <s> <s*>; <s=>  
1030 "<K><O>" <KO>  
1031 translit_end  
1032
```

1033 The "translit\_start" keyword introduces the transliteration section in the LC\_CTYPE category.

1034  
 1035 The "include" keyword specifies that the FDCC-set "de\_DE" is copied and that the repertoiremap "de\_repmap" is  
 1036 used to define the symbolic character names in the FDCC-set "de\_DE".

1037  
 1038 The "default\_missing" keyword introduces the character sequence "<?>" as the string to transform into for input  
 1039 characters that cannot be transformed into other strings, because no transliteration statement is applicable to the

1040 character.

1041

1042 The next 3 lines are transliteration statements.

1043

1044 The first transliteration statement defines a number of transliterations for the LATIN LETTER AE, including into  
1045 LATIN LETTER A WITH DIAERESIS, GREEK LETTER EPSILON, the two Latin letters A and E, and finally  
1046 the LATIN LETTER E.

1047

1048 The second transliteration statement defines transliteration of the LATIN LETTER S into GREEK LETTER  
1049 SIGMA, and CYRILLIC LETTER ES.

1050

1051 The third transliteration statement transliterates the two Latin letters K and O into the Japanese Hiragana character  
1052 KO.

1053

1054 The transliteration sections is terminated via the "translit\_end" keyword in the above example.

1055

### 1056 4.2.3 "i18n" LC\_CTYPE category

1057

1058 The "i18n" FDCC-set for the LC\_CTYPE is defined as follows:

1059

```

1060 LC_CTYPE
1061 % The following is the 14652 i18n fdcc-set LC_CTYPE category.
1062 % It covers ISO/IEC 10646-1 including Cor.1 and AMD 1 thru 9
1063 % The "upper" class reflects the uppercase characters of class "alpha"
1064 upper /
1065 % TABLE 1 BASIC LATIN
1066 <U0041>..<U005A>;/
1067 % TABLE 2 LATIN-1 SUPPLEMENT
1068 <U00C0>..<U00D6>;<U00D8>..<U00DE>;/
1069 % TABLE 3 LATIN EXTENDED-A
1070 <U0100>..(2)..<U0136>;/
1071 <U0139>..(2)..<U0147>;/
1072 <U014A>..(2)..<U0178>;/
1073 <U0179>..(2)..<U017D>;/
1074 % TABLE 4 LATIN EXTENDED-B
1075 <U0181>;<U0182>..(2)..<U0186>;<U0187>;/
1076 <U0189>..<U018B>;<U018E>..<U0191>;<U0193>;<U0194>;/
1077 <U0196>..<U0198>;<U019C>;<U019D>;<U019F>;/
1078 <U01A0>..(2)..<U01A4>;/
1079 <U01A7>;<U01A9>;<U01AC>;<U01AE>;<U01AF>;<U01B1>..<U01B3>;/
1080 <U01B5>;<U01B7>;<U01B8>;<U01BC>;<U01C4>;<U01C5>;<U01C7>;<U01C8>;/
1081 <U01CA>;<U01CB>;/
1082 <U01CD>..(2)..<U01DB>;/
1083 <U01DE>..(2)..<U01EE>;/
1084 <U01F1>;<U01F2>;<U01F4>;<U01FA>..(2)..<U01FE>/
1085 % TABLE 5 LATIN EXTENDED-B
1086 <U0200>..(2)..<U0216>;/
1087 % TABLE 6 IPA EXTENSIONS
1088 <U0262>;<U026A>;<U0274>;<U0276>;/
1089 <U0280>;<U0281>;<U028F>;<U0299>;<U029B>;<U029C>;<U029F>;/
1090 % TABLE 9 BASIC GREEK
1091 <U0386>;<U0388>..<U038A>;<U038C>;<U038E>;<U038F>;<U0391>..<U03A1>;/
1092 <U03A3>..<U03AB>;/
1093 % TABLE 10 GREEK SYMBOLS AND COPTIC
1094 <U03E3>..(2)..<U3EE>;/
1095 % TABLE 11 CYRILLIC
1096 <U0401>..<U040C>;<U040E>..<U042F>;<U0460>..(2)..<U047E>;/
1097 % TABLE 12 CYRILLIC
1098 <U0480>;<U0490>..(2)..<U04BE>;<U04C1>;<U04C3>;<U04C7>;<U04CB>;/
1099 <U04D0>..(2)..<U04EA>;<U04EE>..(2)..<U04F4>;<U04F8>;/
1100 % TABLE 13 ARMENIAN
1101 <U0531>..<U0556>;/
1102 % TABLE 28 GEORGIAN
1103 <U10A0>..<U10C5>;/
1104 % TABLE 31 LATIN EXTENDED ADDITIONAL
1105 <U1E00>..(2)..<U1E7E>;/
1106 % TABLE 32 LATIN EXTENDED ADDITIONAL
1107 <U1E80>..(2)..<U1E94>;/
1108 <U1EA0>..(2)..<U1EF8>;/
1109 % TABLE 33 GREEK EXTENDED
1110 <U1F08>..<U1F0F>;<U1F18>..<U1F1D>;<U1F28>..<U1F2F>;<U1F38>..<U1F3F>;/
1111 <U1F48>..<U1F4D>;<U1F59>..(2)..<U1F5F>;<U1F68>..<U1F6F>;/

```

```

1112 % TABLE 34 GREEK EXTENDED
1113 <U1F88>..<U1F8F>;<U1F98>..<U1F9F>;<U1FA8>..<U1FAF>;<U1FB8>..<U1FBC>;/
1114 <U1FC8>..<U1FCC>;<U1FD8>..<U1FDB>;<U1FE8>..<U1FEC>;<U1FF8>..<U1FFC>
1115 % TABLE 28 GEORGIAN is not addressed as the letters does not have
1116 % a uppercase/lowercase relation
1117 %
1118 % The "lower" class reflects the lowercase characters of class "alpha"
1119 lower /
1120 % TABLE 1 BASIC LATIN
1121 <U0061>..<U007A>;/
1122 % TABLE 2 LATIN-1 SUPPLEMENT
1123 <U00DF>..<U00F6>;<U00F8>..<U00FF>;/
1124 % TABLE 3 LATIN EXTENDED-A
1125 <U0101>..(2)..<U0148>;<U0149>..(2)..<U0177>;<U017A>..(2)..<U017E>;<U017F>;/
1126 % TABLE 4 LATIN EXTENDED-B
1127 <U0180>;<U0183>;<U0185>;<U0188>;<U018C>;<U018D>;<U0192>;<U0195>;/
1128 <U0199>..<U019B>;<U019E>;<U01A1>;<U01A3>;<U01A5>;<U01A8>;<U01AB>;<U01AD>;/
1129 <U01B0>;<U01B4>;<U01B6>;<U01B9>;<U01BA>;<U01BD>;<U01C5>;<U01C6>;/
1130 <U01C8>;<U01C9>;<U01CB>;<U01CC>..(2)..<U01DC>;/
1131 <U01DD>;..(2)..<U01F2>;<U01F3>;<U01F5>;<U01FB>;<U01FD>;<U01FF>;/
1132 % TABLE 5 LATIN EXTENDED-B
1133 <U0201>..(2)..<U0217>;/
1134 % TABLE 6 IPA EXTENSIONS
1135 <U0250>..<U0293>;<U0299>..<U02A0>;<U02A3>..<U02A8>;/
1136 % TABLE 9 BASIC GREEK
1137 <U0390>;<U03AC>..<U03CE>;/
1138 % TABLE 10 GREEK SYMBOLS AND COPTIC
1139 <U03E3>..(2)..<U03EF>/
1140 % TABLE 11 CYRILLIC
1141 <U0430>..<U044F>;<U0451>..<U045C>;<U045E>;<U045F>;<U0460>..(2)..<U047F>;/
1142 % TABLE 12 CYRILLIC
1143 <U04801>;<U0490>..(2)..<U04BF>;<U04C2>;<U04C4>;<U04C8>;<U04CC>;/
1144 <U04D1>..(2)..<U04EB>;<U04EF>..(2)..<U04F5>;<U04F9>;/
1145 % TABLE 13 ARMENIAN
1146 <U0561>..<U0587>;/
1147 % TABLE 28 GEORGIAN
1148 <U10D0>..<U10F6>;/
1149 % TABLE 31 and 32 LATIN EXTENDED ADDITIONAL
1150 <U1E01>..(2)..<U1E95>;<U1EA1>..(2)..<U1EF9>;/
1151 % TABLE 33 and 34 GREEK EXTENDED
1152 <U1F08>..<U1F0F>;<U1F18>..<U1F1D>;<U1F28>..<U1F2F>;<U1F38>..<U1F3F>;/
1153 <U1F48>..<U1F4D>;<U1F59>..(2)..<U1F5F>;<U1F68>..<U1F6F>;/
1154 % TABLE 34 GREEK EXTENDED
1155 <U1F00>..<U1F07>;<U1F10>..<U1F15>;<U1F20>..<U1F27>;<U1F30>..<U1F37>;/
1156 <U1F40>..<U1F45>;<U1F50>..<U1F57>;<U1F60>..<U1F67>;<U1F70>..<U1F7D>;/
1157 <U1F80>..<U1F87>;<U1F90>..<U1F97>;<U1FA0>..<U1FA7>;<U1FB0>..<U1FB4>;/
1158 <U1FB6>;<U1FB7>;<U1FC2>..<U1FC4>;<U1FC6>;<U1FC7>;<U1FD0>..<U1FD3>;/
1159 <U1FD6>;<U1FD7>;<U1FE0>..<U1FE7>;<U1FF2>..<U1FF4>;<U1FF6>;<U1FF7>;/
1160 % TABLE 35 SUPERSCRIPTS AND SUBSCRIPTS, CURRENCY SYMBOLS
1161 <U207F>
1162 %
1163 % The "alpha" class of the "i18n" FDCC-set is reflecting
1164 % the recommendations in TR 10176 annex A
1165 alpha /
1166 % TABLE 1 BASIC LATIN
1167 <U0041>..<U005A>;<U0061>..<U007A>;/
1168 % TABLE 2 LATIN-1 SUPPLEMENT
1169 <U00AA>;<U00BA>;<U00C0>..<U00D6>;<U00D8>..<U00F6>;<U00F8>..<U00FF>;/
1170 % TABLE 3 LATIN EXTENDED-A
1171 <U0100>..<U017F>;/
1172 % TABLE 4 and 5 LATIN EXTENDED-B
1173 <U0180>..<U01F5>;<U01FA>..<U0217>;/
1174 % TABLE 6 IPA EXTENSIONS
1175 <U0250>..<U02A8>;/
1176 % TABLE 31 and 32 LATIN EXTENDED ADDITIONAL
1177 <U1E00>..<U1E9B>;<U1EA0>..<U1EF9>;/
1178 % TABLE 35 SUPERSCRIPTS AND SUBSCRIPTS, CURRENCY SYMBOLS
1179 <U207F>;/
1180 % TABLE 9 BASIC GREEK
1181 <U0386>;<U0388>..<U038A>;<U038C>;<U038E>..<U03A1>;<U03A3>..<U03CE>;/
1182 % TABLE 10 GREEK SYMBOLS AND COPTIC
1183 <U03D0>..<U03D6>;<U03DA>;<U03DC>;<U03DE>;<U03E0>;<U03E2>..<U03F3>;/
1184 % TABLE 33 and 34 GREEK EXTENDED
1185 <U1F00>..<U1F15>;<U1F18>..<U1F1D>;<U1F20>..<U1F45>;<U1F48>..<U1F4D>;/
1186 <U1F50>..<U1F57>;<U1F59>;<U1F5B>;<U1F5D>;<U1F5F>..<U1F7D>;/
1187 <U1F80>..<U1FB4>;<U1FB6>..<U1FBC>;<U1FC2>..<U1FC4>;<U1FC6>..<U1FCC>;/
1188 <U1FD0>..<U1FD3>;<U1FD6>..<U1FDB>;<U1FE0>..<U1FEC>;<U1FF2>..<U1FF4>;/

```



1189 <U1FF6>..<U1FFC>;/  
 1190 % TABLE 11 and 12 CYRILLIC  
 1191 <U0401>..<U040C>;<U040E>..<U044F>;<U0451>..<U045C>;<U045E>..<U0481>;/  
 1192 <U0490>..<U04C4>;<U04C7>..<U04C8>;<U04CB>..<U04CC>;<U04D0>..<U04EB>;/  
 1193 <U04EE>..<U04F5>;<U04F8>..<U04F9>;/  
 1194 % TABLE 13 ARMENIAN  
 1195 <U0531>..<U0556>;<U0561>..<U0587>;/  
 1196 % TABLE 14 HEBREW  
 1197 <U05B0>..<U05B9>;<U05BB>..<U05BD>;<U05BF>;<U05C1>..<U05C2>;/  
 1198 <U05D0>..<U05EA>;<U05F0>..<U05F2>;/  
 1199 % TABLE 15 and 16 ARABIC  
 1200 <U0621>..<U063A>;<U0640>..<U0652>;<U0670>..<U06B7>;<U06BA>..<U06BE>;/  
 1201 <U06C0>..<U06CE>;<U06D0>..<U06D3>;<U06D5>..<U06DC>;<U06E5>..<U06E8>;/  
 1202 <U06EA>..<U06ED>;/  
 1203 % TABLE 17 DEVANAGARI  
 1204 <U0901>..<U0903>;<U0905>..<U0939>;<U093E>..<U094D>;<U0950>..<U0952>;/  
 1205 <U0958>..<U0963>;/  
 1206 % TABLE 18 BENGALI  
 1207 <U0981>..<U0983>;<U0985>..<U098C>;<U098F>..<U0990>;/  
 1208 <U0993>..<U09A8>;<U09AA>..<U09B0>;<U09B2>;<U09B6>..<U09B9>;/  
 1209 <U09BE>..<U09C4>;<U09C7>..<U09C8>;<U09CB>..<U09CD>;<U09DC>..<U09DD>;/  
 1210 <U09DF>..<U09E3>;<U09F0>..<U09F1>;/  
 1211 % TABLE 19 GURMUKHI  
 1212 <U0A02>;<U0A05>..<U0A0A>;<U0A0F>..<U0A10>;<U0A13>..<U0A28>;/  
 1213 <U0A2A>..<U0A30>;<U0A32>..<U0A33>;<U0A35>..<U0A36>;<U0A38>..<U0A39>;/  
 1214 <U0A3E>..<U0A42>;<U0A47>..<U0A48>;<U0A4B>..<U0A4D>;<U0A59>..<U0A5C>;/  
 1215 <U0A5E>;<U0A74>;/  
 1216 % TABLE 20 GUJARATI  
 1217 <U0A81>..<U0A83>;<U0A85>..<U0A8B>;<U0A8D>;<U0A8F>..<U0A91>;/  
 1218 <U0A93>..<U0AA8>;<U0AAA>..<U0AB0>;<U0AB2>..<U0AB3>;<U0AB5>..<U0AB9>;/  
 1219 <U0ABD>..<U0AC5>;<U0AC7>..<U0AC9>;<U0ACB>..<U0ACD>;<U0AD0>;<U0AE0>;/  
 1220 % TABLE 21 ORIYA  
 1221 <U0B01>..<U0B03>;<U0B05>..<U0B0C>;<U0B0F>..<U0B10>;<U0B13>..<U0B28>;/  
 1222 <U0B2A>..<U0B30>;<U0B32>..<U0B33>;<U0B36>..<U0B3E>..<U0B43>;/  
 1223 <U0B47>..<U0B48>;<U0B4B>..<U0B4D>;<U0B5C>..<U0B5D>;<U0B5F>..<U0B61>;/  
 1224 % TABLE 22 TAMIL  
 1225 <U0B82>..<U0B83>;<U0B85>..<U0B8A>;<U0B8E>..<U0B90>;<U0B92>..<U0B95>;/  
 1226 <U0B99>..<U0B9A>;<U0B9C>;<U0B9E>..<U0B9F>;<U0BA3>..<U0BA4>;/  
 1227 <U0BA8>..<U0BAA>;<U0BAE>..<U0BB5>;<U0BB7>..<U0BB9>;<U0BBE>..<U0BC2>;/  
 1228 <U0BC6>..<U0BC8>;<U0BCA>..<U0BCD>;/  
 1229 % TABLE 23 TELUGU  
 1230 <U0C01>..<U0C03>;<U0C05>..<U0C0C>;<U0C0E>..<U0C10>;<U0C12>..<U0C28>;/  
 1231 <U0C2A>..<U0C33>;<U0C35>..<U0C39>;<U0C3E>..<U0C44>;<U0C46>..<U0C48>;/  
 1232 <U0C4A>..<U0C4D>;<U0C60>..<U0C61>;/  
 1233 % TABLE 24 KANNADA  
 1234 <U0C82>..<U0C83>;<U0C85>..<U0C8C>;<U0C8E>..<U0C90>;<U0C92>..<U0CA8>;/  
 1235 <U0CAA>..<U0CB3>;<U0CB5>..<U0CB9>;<U0CBE>..<U0CC4>;<U0CC6>..<U0CC8>;/  
 1236 <U0CCA>..<U0CCD>;<U0CDE>;<U0CE0>..<U0CE1>;/  
 1237 % TABLE 25 MALAYALAM  
 1238 <U0D02>..<U0D03>;<U0D05>..<U0D0C>;<U0D0E>..<U0D10>;<U0D12>..<U0D28>;/  
 1239 <U0D2A>..<U0D39>;<U0D3E>..<U0D43>;<U0D46>..<U0D48>;<U0D4A>..<U0D4D>;/  
 1240 <U0D60>..<U0D61>;/  
 1241 % TABLE 26 THAI  
 1242 <U0E01>..<U0E3A>;<U0E40>..<U0E4E>;<U0E50>..<U0E59>;/  
 1243 % TABLE 27 LAO  
 1244 <U0E81>..<U0E82>;<U0E84>;<U0E87>..<U0E88>;<U0E8A>;<U0E8D>;/  
 1245 <U0E94>..<U0E97>;<U0E99>..<U0E9F>;<U0EA1>..<U0EA3>;<U0EA5>;<U0EA7>;/  
 1246 <U0EAA>..<U0EAB>;<U0EAD>..<U0EAE>;<U0EB0>..<U0EB9>;<U0EBB>..<U0EBD>;/  
 1247 <U0EC0>..<U0EC4>;<U0EC6>;<U0EC8>..<U0ECD>;<U0EDC>..<U0EDD>;/  
 1248 % TIBETAN Amendment 6  
 1249 <U0F00>;<U0F18>..<U0F19>;<U0F35>;<U0F37>;<U0F39>;<U0F40>..<U0F47>;/  
 1250 <U0F49>..<U0F69>;/  
 1251 <U0F71>..<U0F84>;<U0F86>..<U0F8B>;<U0F90>..<U0F95>;<U0F97>;/  
 1252 <U0F99>..<U0FAD>;<U0FB1>..<U0FB7>;<U0FB9>;/  
 1253 % TABLE 28 GEORGIAN  
 1254 <U10A0>..<U10C5>;<U10D0>..<U10F6>;/  
 1255 % TABLE 50 HIRAGANA  
 1256 <U3041>..<U3093>;<U309B>..<U309C>;/  
 1257 % TABLE 51 KATAKANA  
 1258 <U30A1>..<U30F6>;<U30FB>..<U30FC>;/  
 1259 % TABLE 52 BOPOMOFO  
 1260 <U3105>..<U312C>;/  
 1261 % CJK unified ideographs  
 1262 <U4E01>..<U9FA5>;/  
 1263 % HANGUL amendment 5  
 1264 <UAC00>..<UD7A3>;/  
 1265 % Miscellaneous  
 1266 <U00B5>;<U00B7>;<U02B0>..<U02B8>;<U02BB>;<U02BD>..<U02C1>;/  
 1267 <U02D0>..<U02D1>;<U02E0>..<U02E4>;<U037A>;<U0559>;<U093D>;<U0B3D>;/

```

1268 <U1FBE>;<U203F>..<U2040>;<U2102>;<U2107>;<U210A>..<U2113>;<U2115>;/
1269 <U2118>..<U211D>;<U2124>;<U2126>;<U2128>;<U212A>..<U2131>;/
1270 <U2133>..<U2138>;<U2160>..<U2182>;<U3005>..<U3006>;<U3021>..<U3029>
1271 %
1272 % The "digit" class of the "i18n" FDCC-set is reflecting
1273 % the recommendations in TR 10176 annex A
1274 digit /
1275 <U0030>..<U0039>;<U0660>..<U0669>;<U06F0>..<U06F9>;<U0966>..<U096F>;/
1276 <U09E6>..<U09EF>;<U0A66>..<U0A6F>;<U0AE6>..<U0AEF>;<U0B66>..<U0B6F>;/
1277 <0>;<U0BE7>..<U0BEF>;<U0C66>..<U0C6F>;<U0CE6>..<U0CEF>;<U0D66>..<U0D6F>;/
1278 <U0E50>..<U0E59>;<U0ED0>..<U0ED9>;<U0F20>..<U0F29>
1279 %
1280 outdigit <U0030>..<U0039>
1281 %
1282 space <U0008>;<U000A>..<U000D>;<U0020>;<U2000>..<U2006>;/
1283 <U2008>..<U200B>;<U3000>
1284 %
1285 cntrl <U0000>..<U001F>;<U0077>..<U009F>
1286 %
1287 punct /
1288 <U0021>..<U002F>;<U003A>..<U0040>;<U005B>..<U0060>;/
1289 <U007B>..<U007E>;<U00A0>..<U00A9>;<U00AB>..<U00B9>;<U00BB>..<U00BF>;/
1290 <U00D7>;<U00F7>;<U02C7>;<U02D8>..<U02DD>;/
1291 <U037E>;<U0482>;<U055A>..<U055F>;<U0589>;<U05BE>;<U05C0>;<U05C3>;/
1292 <U05F3>;<U05F4>;<U060C>;<U061B>;<U061F>;<U0640>;<U064B>..<U0652>;/
1293 <U066A>..<U066D>;<U06D4>;<U06DD>..<U06E1>;<U06E9>..<U06EC>;<U10FB>;/
1294 <U2010>..<U2029>;<U2030>..<U2046>;<U20A0>..<U20AA>;<U2100>..<U210B>;/
1295 <U210D>..<U2110>;<U2112>..<U211B>;<U211D>..<U2127>;<U212A>..<U212C>;/
1296 <U212E>..<U2138>;<U2200>..<U22F1>;<U2300>;<U2302>..<U237A>;<U2400>..<U2424>;/
1297 <U2440>..<U244A>;<U2580>..<U2595>;<U25A0>..<U25EF>;<U2600>..<U2613>;/
1298 <U261A>..<U266F>;<U2701>..<U2704>;<U2706>..<U2709>;<U270C>..<U2727>;/
1299 <U2729>..<U274B>;<U274D>;<U274F>..<U2752>;<U2756>;<U2758>..<U275E>;/
1300 <U2761>..<U2767>;<U3000>..<U3020>;<U3030>;<U3036>;<U3037>;<U303F>;<U3164>;/
1301 <U3190>..<U319F>;<U3200>..<U321C>;<U3220>..<U3243>;<U3260>..<U327B>;/
1302 <U327F>..<U32B0>;<U32C0>..<U32CB>;<U32D0>..<U32FE>;<U3300>..<U3376>;/
1303 <U337B>..<U33DD>;<U33E0>..<U33FE>;<UFD3E>;<UFD3F>;<UFE49>..<UFE52>;/
1304 <UFE54>..<UFE66>;<UFE68>..<UFE6B>;<UFEFF>;<UFF01>..<UFF0F>;<UFF1A>..<UFF20>;/
1305 <UFF3B>..<UFF40>;<UFF5B>..<UFF5E>;<UFF61>..<UFF65>;<UFF70>;<UFF9E>..<UFFA0>;/
1306 <UFFE0>..<UFFE6>;<UFFE8>..<UFFEE>;<UFFFD>
1307 %
1308 graph /
1309 <U0021>..<U007E>;<U00A0>..<U01F5>;<U01FA>..<U0217>;/
1310 <U0250>..<U02A8>;<U02B0>..<U02DE>;<U02E0>..<U02E9>;<U0300>..<U0345>;/
1311 <U0360>;<U0361>;<U0374>;<U0375>;<U037A>;<U037E>;<U0384>..<U038A>;<U038C>;/
1312 <U038E>..<U03A1>;<U03A3>..<U03CE>;<U03D0>..<U03D6>;<U03DA>;<U03DC>;<U03DE>;/
1313 <U03E0>;<U03E2>..<U03F3>;<U0401>..<U040C>;<U040E>..<U044F>;/
1314 <U0451>..<U045C>;<U045E>..<U0486>;<U0490>..<U04C4>;<U04C7>;<U04C8>;/
1315 <U04CB>;<U04CC>;<U04D0>..<U04EB>;<U04EE>..<U04F5>;<U04F8>;<U04F9>;/
1316 <U0531>..<U0556>;<U0559>..<U055F>;<U0561>..<U0587>;<U0589>;/
1317 <U0591>..<U05A1>;<U05A3>..<U05AF>;<U05B0>..<U05B9>;/
1318 <U05BB>..<U05C4>;<U05D0>..<U05EA>;<U05F0>..<U05F4>;<U060C>;<U061B>;<U061F>;/
1319 <U0621>..<U063A>;<U0640>..<U0652>;<U0660>..<U066D>;<U0670>..<U06B7>;/
1320 <U06BA>..<U06BE>;<U06C0>..<U06CE>;<U06D0>..<U06ED>;<U06F0>..<U06F9>;/
1321 <U0901>..<U0903>;<U0905>..<U0939>;<U093C>..<U094D>;<U0950>..<U0954>;/
1322 <U0958>..<U0970>;<U0981>..<U0983>;<U0985>..<U098C>;<U098F>;<U0990>;/
1323 <U0993>..<U09A8>;<U09AA>..<U09B0>;<U09B2>;<U09B6>..<U09B9>;<U09BC>;/
1324 <U09BE>..<U09C4>;<U09C7>;<U09C8>;<U09CB>..<U09CD>;<U09D7>;<U09DC>;<U09DD>;/
1325 <U09DF>..<U09E3>;<U09E6>..<U09FA>;<U0A02>;<U0A05>..<U0A0A>;<U0A0F>;<U0A10>;/
1326 <U0A13>..<U0A28>;<U0A2A>..<U0A30>;<U0A32>;<U0A33>;<U0A35>;<U0A36>;/
1327 <U0A38>;<U0A39>;<U0A3C>;<U0A3E>..<U0A42>;<U0A47>;<U0A48>;<U0A4B>..<U0A4D>;/
1328 <U0A59>..<U0A5C>;<U0A5E>;<U0A66>..<U0A74>;<U0A81>..<U0A83>;<U0A85>..<U0A8B>;/
1329 <U0A8D>;<U0A8F>..<U0A91>;<U0A93>..<U0A98>;<U0AAA>..<U0AB0>;/
1330 <U0AB2>;<U0AB3>;<U0AB5>..<U0AB9>;<U0ABC>..<U0AC5>;<U0AC7>..<U0AC9>;/
1331 <U0ACB>..<U0ACD>;<U0AD0>;<U0AE0>;<U0AE6>..<U0AEF>;<U0B01>..<U0B03>;/
1332 <U0B05>..<U0B0C>;<U0B0F>;<U0B10>;<U0B13>..<U0B28>;<U0B2A>..<U0B30>;/
1333 <U0B32>;<U0B33>;<U0B36>..<U0B39>;<U0B3C>..<U0B43>;<U0B47>;<U0B48>;/
1334 <U0B4B>..<U0B4D>;<U0B56>;<U0B57>;<U0B5C>;<U0B5D>;<U0B5F>..<U0B61>;/
1335 <U0B66>..<U0B70>;<U0B82>;<U0B83>;<U0B85>..<U0B8A>;<U0B8E>..<U0B90>;/
1336 <U0B92>..<U0B95>;<U0B99>;<U0B9A>;<U0B9C>;<U0B9E>;<U0B9F>;<U0BA3>;<U0BA4>;/
1337 <U0BA8>..<U0BAA>;<U0BAE>..<U0BB5>;<U0BB7>..<U0BB9>;<U0BBE>..<U0BC2>;/
1338 <U0BC6>..<U0BC8>;<U0BCA>..<U0BCD>;<U0BD7>;<U0BE7>..<U0BF2>;<U0C01>..<U0C03>;/
1339 <U0C05>..<U0C0C>;<U0C0E>..<U0C10>;<U0C12>..<U0C28>;<U0C2A>..<U0C33>;/
1340 <U0C35>..<U0C39>;<U0C3E>..<U0C44>;<U0C46>..<U0C48>;<U0C4A>..<U0C4D>;/
1341 <U0C55>;<U0C56>;<U0C60>;<U0C61>;<U0C66>..<U0C6F>;<U0C82>;<U0C83>;/
1342 <U0C85>..<U0C8C>;<U0C8E>..<U0C90>;<U0C92>..<U0CA8>;<U0CAA>..<U0CB3>;/
1343 <U0CB5>..<U0CB9>;<U0CBE>..<U0CC4>;<U0CC6>..<U0CC8>;<U0CCA>..<U0CCD>;/
1344 <U0CD5>;<U0CD6>;<U0CDE>;<U0CE0>;<U0CE1>;<U0CE6>..<U0CEF>;<U0D02>;<U0D03>;/

```

```

1345 <U0D05>..<U0D0C>;<U0D0E>..<U0D10>;<U0D12>..<U0D28>;<U0D2A>..<U0D39>;/
1346 <U0D3E>..<U0D43>;<U0D46>..<U0D48>;<U0D4A>..<U0D4D>;<U0D57>;<U0D60>;<U0D61>;/
1347 <U0D66>..<U0D6F>;<U0E01>..<U0E3A>;<U0E3F>..<U0E5B>;<U0E81>;<U0E82>;<U0E84>;/
1348 <U0E87>;<U0E88>;<U0E8A>;<U0E8D>;<U0E94>..<U0E97>;<U0E99>..<U0E9F>;/
1349 <U0EA1>..<U0EA3>;<U0EA5>;<U0EA7>;<U0EAA>;<U0EAB>;<U0EAD>..<U0EB9>;/
1350 <U0EBB>..<U0EBD>;<U0EC0>..<U0EC4>;<U0EC6>;<U0EC8>..<U0ECD>;<U0ED0>..<U0ED9>;/
1351 <U0EDC>;<U0EDD>;/
1352 <U0F00>..<U0F47>;<U0F49>..<U0F69>;<U0F71>..<U0F7F>;/
1353 <U10A0>..<U10C5>;<U10D0>..<U10F6>;<U10FB>;<U1100>..<U1159>;/
1354 <U115F>..<U11A2>;<U11A8>..<U11F9>;<U1E00>..<U1E9B>;<U1EA0>..<U1EF9>;/
1355 <U1F00>..<U1F15>;<U1F18>..<U1F1D>;<U1F20>..<U1F45>;<U1F48>..<U1F4D>;/
1356 <U1F50>..<U1F57>;<U1F59>;<U1F5B>;<U1F5D>;<U1F5F>..<U1F7D>;<U1F80>..<U1FB4>;/
1357 <U1FB6>..<U1FC4>;<U1FC6>..<U1FD3>;<U1FD6>..<U1FDB>;<U1FDD>..<U1FEF>;/
1358 <U1FF2>..<U1FF4>;<U1FF6>..<U1FFE>;<U2000>..<U202E>;<U2030>..<U2046>;/
1359 <U206A>..<U2070>;<U2074>..<U208E>;<U20A0>..<U20AB>;<U20D0>..<U20E1>;/
1360 <U2100>..<U2138>;<U2153>..<U2182>;<U2190>..<U21EA>;<U2200>..<U22F1>;<U2300>;/
1361 <U2302>..<U237A>;<U2400>..<U2424>;<U2440>..<U244A>;<U2460>..<U24EA>;/
1362 <U2500>..<U2595>;<U25A0>..<U25EF>;<U2600>..<U2613>;<U261A>..<U266F>;/
1363 <U2701>..<U2704>;<U2706>..<U2709>;<U270C>..<U2727>;<U2729>..<U274B>;<U274D>;/
1364 <U274F>..<U2752>;<U2756>;<U2758>..<U275E>;<U2761>..<U2767>;<U2776>..<U2794>;/
1365 <U2798>..<U27AF>;<U27B1>..<U27BE>;<U3000>..<U3037>;<U303F>;<U3041>..<U3094>;/
1366 <U3099>..<U309E>;<U30A1>..<U30FE>;<U3105>..<U312C>;<U3131>..<U318E>;/
1367 <U3190>..<U319F>;<U3200>..<U321C>;<U3220>..<U3243>;<U3260>..<U327B>;/
1368 <U327F>..<U32B0>;<U32C0>..<U32CB>;<U32D0>..<U32FE>;<U3300>..<U3376>;/
1369 <U337B>..<U33DD>;<U33E0>..<U33FE>;<UFB00>..<UFB06>;<UFB13>..<UFB17>;/
1370 <UFB1E>..<UFB36>;<UFB38>..<UFB3C>;<UFB3E>;<UFB40>;<UFB41>;<UFB43>;<UFB44>;/
1371 <UFB46>..<UFB61>;<UFB63>..<UFD3F>;<UFD50>..<UFD8F>;<UFD92>..<UFD7C>;/
1372 <UFD7F>..<UFD7E>;<UFE20>..<UFE23>;<UFE30>..<UFE44>;<UFE49>..<UFE52>;/
1373 <UFE54>..<UFE66>;<UFE68>..<UFE6B>;<UFE70>..<UFE72>;<UFE74>;<UFE76>..<UFEFC>;/
1374 <UFEFF>;<UFF01>..<UFF5E>;<UFF61>..<UFFBE>;<UFFC2>..<UFFC7>;/
1375 <UFFCA>..<UFFCD>;<UFFD2>..<UFFD7>;<UFFDA>..<UFFDC>;<UFFE0>..<UFFE6>;/
1376 <UFFE8>..<UFFEE>;<UFFFD>
1377 %
1378 % "print" is by default "graph", and the <space> character
1379 %
1380 % xdigit <U0030>..<U0039>;<U0041>..<U0046>;<U0061>..<U0066>
1381 %
1382 % blank <U0008>;<U0020>;<U2000>..<U2006>;<U2008>..<U200B>;<U3000>
1383 %
1384 % toupper /
1385 (<U0061>,<U0041>);(<U0062>,<U0042>);(<U0063>,<U0043>);(<U0064>,<U0044>);/
1386 (<U0065>,<U0045>);(<U0066>,<U0046>);(<U0067>,<U0047>);(<U0068>,<U0048>);/
1387 (<U0069>,<U0049>);(<U006A>,<U004A>);(<U006B>,<U004B>);(<U006C>,<U004C>);/
1388 (<U006D>,<U004D>);(<U006E>,<U004E>);(<U006F>,<U004F>);(<U0070>,<U0050>);/
1389 (<U0071>,<U0051>);(<U0072>,<U0052>);(<U0073>,<U0053>);(<U0074>,<U0054>);/
1390 (<U0075>,<U0055>);(<U0076>,<U0056>);(<U0077>,<U0057>);(<U0078>,<U0058>);/
1391 (<U0079>,<U0059>);(<U007A>,<U005A>);(<U00E0>,<U00C0>);(<U00E1>,<U00C1>);/
1392 (<U00E2>,<U00C2>);(<U00E3>,<U00C3>);(<U00E4>,<U00C4>);(<U00E5>,<U00C5>);/
1393 (<U00E6>,<U00C6>);(<U00E7>,<U00C7>);(<U00E8>,<U00C8>);(<U00E9>,<U00C9>);/
1394 (<U00EA>,<U00CA>);(<U00EB>,<U00CB>);(<U00EC>,<U00CC>);(<U00ED>,<U00CD>);/
1395 (<U00EE>,<U00CE>);(<U00EF>,<U00CF>);(<U00F0>,<U00D0>);(<U00F1>,<U00D1>);/
1396 (<U00F2>,<U00D2>);(<U00F3>,<U00D3>);(<U00F4>,<U00D4>);(<U00F5>,<U00D5>);/
1397 (<U00F6>,<U00D6>);(<U00F7>,<U00D7>);(<U00F8>,<U00D8>);(<U00F9>,<U00D9>);/
1398 (<U00FA>,<U00DA>);(<U00FB>,<U00DB>);(<U00FC>,<U00DC>);(<U00FD>,<U00DD>);/
1399 (<U00FE>,<U00DE>);(<U00FF>,<U00DF>);(<U0101>,<U0100>);(<U0103>,<U0102>);/
1400 (<U0105>,<U0104>);(<U0107>,<U0106>);(<U0109>,<U0108>);(<U010B>,<U010A>);/
1401 (<U010D>,<U010C>);(<U010F>,<U010E>);(<U0111>,<U0110>);(<U0113>,<U0112>);/
1402 (<U0115>,<U0114>);(<U0117>,<U0116>);(<U0119>,<U0118>);(<U011B>,<U011A>);/
1403 (<U011D>,<U011C>);(<U011F>,<U011E>);(<U0121>,<U0120>);(<U0123>,<U0122>);/
1404 (<U0125>,<U0124>);(<U0127>,<U0126>);(<U0129>,<U0128>);(<U012B>,<U012A>);/
1405 (<U012D>,<U012C>);(<U012F>,<U012E>);(<U0133>,<U0132>);(<U0135>,<U0134>);/
1406 (<U0137>,<U0136>);(<U013A>,<U0139>);(<U013C>,<U013B>);(<U013E>,<U013D>);/
1407 (<U0140>,<U013F>);(<U0142>,<U0141>);(<U0144>,<U0143>);(<U0146>,<U0145>);/
1408 (<U0148>,<U0147>);(<U014B>,<U014A>);(<U014D>,<U014C>);(<U014F>,<U014E>);/
1409 (<U0151>,<U0150>);(<U0153>,<U0152>);(<U0155>,<U0154>);(<U0157>,<U0156>);/
1410 (<U0159>,<U0158>);(<U015B>,<U015A>);(<U015D>,<U015C>);(<U015F>,<U015E>);/
1411 (<U0161>,<U0160>);(<U0163>,<U0162>);(<U0165>,<U0164>);(<U0167>,<U0166>);/
1412 (<U0169>,<U0168>);(<U016B>,<U016A>);(<U016D>,<U016C>);(<U016F>,<U016E>);/
1413 (<U0171>,<U0170>);(<U0173>,<U0172>);(<U0175>,<U0174>);(<U0177>,<U0176>);/
1414 (<U017A>,<U0179>);(<U017C>,<U017B>);(<U017E>,<U017D>);(<U017F>,<U0053>);/
1415 (<U0183>,<U0182>);(<U0185>,<U0184>);(<U0188>,<U0187>);(<U018C>,<U018B>);/
1416 (<U0192>,<U0191>);(<U0199>,<U0198>);(<U01A1>,<U01A0>);(<U01A3>,<U01A2>);/
1417 (<U01A5>,<U01A4>);(<U01A8>,<U01A7>);(<U01AD>,<U01AC>);(<U01B0>,<U01AF>);/
1418 (<U01B4>,<U01B3>);(<U01B6>,<U01B5>);(<U01B9>,<U01B8>);(<U01BD>,<U01BC>);/
1419 (<U01C5>,<U01C4>);(<U01C6>,<U01C4>);(<U01C8>,<U01C7>);/
1420 (<U01C9>,<U01C7>);(<U01CB>,<U01CA>);(<U01CC>,<U01CA>);/
1421 (<U01CE>,<U01CD>);(<U01D0>,<U01CF>);(<U01D2>,<U01D1>);(<U01D4>,<U01D3>);/
1422 (<U01D6>,<U01D5>);(<U01D8>,<U01D7>);(<U01DA>,<U01D9>);(<U01DC>,<U01DB>);/
1423 (<U01DD>,<U018E>);(<U01DF>,<U01DE>);(<U01E1>,<U01E0>);(<U01E3>,<U01E2>);/

```

1424 (<U01E5>, <U01E4>); (<U01E7>, <U01E6>); (<U01E9>, <U01E8>); (<U01EB>, <U01EA>); /  
 1425 (<U01ED>, <U01EC>); (<U01EF>, <U01EE>); (<U01F2>, <U01F1>); /  
 1426 (<U01F3>, <U01F1>); (<U01F5>, <U01F4>); (<U01FB>, <U01FA>); (<U01FD>, <U01FC>); /  
 1427 (<U01FF>, <U01FE>); (<U0201>, <U0200>); (<U0203>, <U0202>); (<U0205>, <U0204>); /  
 1428 (<U0207>, <U0206>); (<U0209>, <U0208>); (<U020B>, <U020A>); (<U020D>, <U020C>); /  
 1429 (<U020F>, <U020E>); (<U0211>, <U0210>); (<U0213>, <U0212>); (<U0215>, <U0214>); /  
 1430 (<U0217>, <U0216>); (<U0253>, <U0181>); (<U0254>, <U0186>); (<U0256>, <U0189>); /  
 1431 (<U0257>, <U018A>); (<U0259>, <U018F>); (<U025B>, <U0190>); (<U0260>, <U0193>); /  
 1432 (<U0263>, <U0194>); (<U0268>, <U0197>); (<U0269>, <U0196>); (<U026F>, <U019C>); /  
 1433 (<U0272>, <U019D>); (<U0275>, <U019F>); (<U0283>, <U01A9>); (<U0288>, <U01AE>); /  
 1434 (<U028A>, <U01B1>); (<U028B>, <U01B2>); (<U0292>, <U01B7>); (<U03AC>, <U0386>); /  
 1435 (<U03AD>, <U0388>); (<U03AE>, <U0389>); (<U03AF>, <U038A>); (<U03B1>, <U0391>); /  
 1436 (<U03B2>, <U0392>); (<U03B3>, <U0393>); (<U03B4>, <U0394>); (<U03B5>, <U0395>); /  
 1437 (<U03B6>, <U0396>); (<U03B7>, <U0397>); (<U03B8>, <U0398>); (<U03B9>, <U0399>); /  
 1438 (<U03BA>, <U039A>); (<U03BB>, <U039B>); (<U03BC>, <U039C>); (<U03BD>, <U039D>); /  
 1439 (<U03BE>, <U039E>); (<U03BF>, <U039F>); (<U03C0>, <U03A0>); (<U03C1>, <U03A1>); /  
 1440 (<U03C2>, <U03A3>); (<U03C3>, <U03A3>); (<U03C4>, <U03A4>); (<U03C5>, <U03A5>); /  
 1441 (<U03C6>, <U03A6>); (<U03C7>, <U03A7>); (<U03C8>, <U03A8>); (<U03C9>, <U03A9>); /  
 1442 (<U03CA>, <U03AA>); (<U03CB>, <U03AB>); (<U03CC>, <U038C>); (<U03CD>, <U038E>); /  
 1443 (<U03CE>, <U038F>); /  
 1444 (<U03E3>, <U03E2>); (<U03E5>, <U03E4>); (<U03E7>, <U03E6>); (<U03E9>, <U03E8>); /  
 1445 (<U03EB>, <U03EA>); (<U03ED>, <U03EC>); (<U03EF>, <U03EE>); /  
 1446 (<U0430>, <U0410>); (<U0431>, <U0411>); (<U0432>, <U0412>); /  
 1447 (<U0433>, <U0413>); (<U0434>, <U0414>); (<U0435>, <U0415>); (<U0436>, <U0416>); /  
 1448 (<U0437>, <U0417>); (<U0438>, <U0418>); (<U0439>, <U0419>); (<U043A>, <U041A>); /  
 1449 (<U043B>, <U041B>); (<U043C>, <U041C>); (<U043D>, <U041D>); (<U043E>, <U041E>); /  
 1450 (<U043F>, <U041F>); (<U0440>, <U0420>); (<U0441>, <U0421>); (<U0442>, <U0422>); /  
 1451 (<U0443>, <U0423>); (<U0444>, <U0424>); (<U0445>, <U0425>); (<U0446>, <U0426>); /  
 1452 (<U0447>, <U0427>); (<U0448>, <U0428>); (<U0449>, <U0429>); (<U044A>, <U042A>); /  
 1453 (<U044B>, <U042B>); (<U044C>, <U042C>); (<U044D>, <U042D>); (<U044E>, <U042E>); /  
 1454 (<U044F>, <U042F>); (<U0451>, <U0401>); (<U0452>, <U0402>); (<U0453>, <U0403>); /  
 1455 (<U0454>, <U0404>); (<U0455>, <U0405>); (<U0456>, <U0406>); (<U0457>, <U0407>); /  
 1456 (<U0458>, <U0408>); (<U0459>, <U0409>); (<U045A>, <U040A>); (<U045B>, <U040B>); /  
 1457 (<U045C>, <U040C>); (<U045E>, <U040E>); (<U045F>, <U040F>); (<U0461>, <U0460>); /  
 1458 (<U0463>, <U0462>); (<U0465>, <U0464>); (<U0467>, <U0466>); (<U0469>, <U0468>); /  
 1459 (<U046B>, <U046A>); (<U046D>, <U046C>); (<U046F>, <U046E>); (<U0471>, <U0470>); /  
 1460 (<U0473>, <U0472>); (<U0475>, <U0474>); (<U0477>, <U0476>); (<U0479>, <U0478>); /  
 1461 (<U047B>, <U047A>); (<U047D>, <U047C>); (<U047F>, <U047E>); (<U0481>, <U0480>); /  
 1462 (<U0491>, <U0490>); (<U0493>, <U0492>); (<U0495>, <U0494>); (<U0497>, <U0496>); /  
 1463 (<U0499>, <U0498>); (<U049B>, <U049A>); (<U049D>, <U049C>); (<U049F>, <U049E>); /  
 1464 (<U04A1>, <U04A0>); (<U04A3>, <U04A2>); (<U04A5>, <U04A4>); (<U04A7>, <U04A6>); /  
 1465 (<U04A9>, <U04A8>); (<U04AB>, <U04AA>); (<U04AD>, <U04AC>); (<U04AF>, <U04AE>); /  
 1466 (<U04B1>, <U04B0>); (<U04B3>, <U04B2>); (<U04B5>, <U04B4>); (<U04B7>, <U04B6>); /  
 1467 (<U04B9>, <U04B8>); (<U04BB>, <U04BA>); (<U04BD>, <U04BC>); (<U04BF>, <U04BE>); /  
 1468 (<U04C2>, <U04C1>); (<U04C4>, <U04C3>); (<U04C8>, <U04C7>); (<U04CC>, <U04CB>); /  
 1469 (<U04D1>, <U04D0>); (<U04D3>, <U04D2>); (<U04D5>, <U04D4>); (<U04D7>, <U04D6>); /  
 1470 (<U04D9>, <U04D8>); (<U04DB>, <U04DA>); (<U04DD>, <U04DC>); (<U04DF>, <U04DE>); /  
 1471 (<U04E1>, <U04E0>); (<U04E3>, <U04E2>); (<U04E5>, <U04E4>); (<U04E7>, <U04E6>); /  
 1472 (<U04E9>, <U04E8>); (<U04EB>, <U04EA>); (<U04EF>, <U04EE>); (<U04F1>, <U04F0>); /  
 1473 (<U04F3>, <U04F2>); (<U04F5>, <U04F4>); (<U04F9>, <U04F8>); (<U0561>, <U0531>); /  
 1474 (<U0562>, <U0532>); (<U0563>, <U0533>); (<U0564>, <U0534>); (<U0565>, <U0535>); /  
 1475 (<U0566>, <U0536>); (<U0567>, <U0537>); (<U0568>, <U0538>); (<U0569>, <U0539>); /  
 1476 (<U056A>, <U053A>); (<U056B>, <U053B>); (<U056C>, <U053C>); (<U056D>, <U053D>); /  
 1477 (<U056E>, <U053E>); (<U056F>, <U053F>); (<U0570>, <U0540>); (<U0571>, <U0541>); /  
 1478 (<U0572>, <U0542>); (<U0573>, <U0543>); (<U0574>, <U0544>); (<U0575>, <U0545>); /  
 1479 (<U0576>, <U0546>); (<U0577>, <U0547>); (<U0578>, <U0548>); (<U0579>, <U0549>); /  
 1480 (<U057A>, <U054A>); (<U057B>, <U054B>); (<U057C>, <U054C>); (<U057D>, <U054D>); /  
 1481 (<U057E>, <U054E>); (<U057F>, <U054F>); (<U0580>, <U0550>); (<U0581>, <U0551>); /  
 1482 (<U0582>, <U0552>); (<U0583>, <U0553>); (<U0584>, <U0554>); (<U0585>, <U0555>); /  
 1483 (<U0586>, <U0556>); /  
 1484 (<U10D0>, <U10A0>); (<U10D1>, <U10A1>); (<U10D2>, <U10A2>); (<U10D3>, <U10A3>); /  
 1485 (<U10D4>, <U10A4>); (<U10D5>, <U10A5>); (<U10D6>, <U10A6>); (<U10D7>, <U10A7>); /  
 1486 (<U10D8>, <U10A8>); (<U10D9>, <U10A9>); (<U10DA>, <U10AA>); (<U10DB>, <U10AB>); /  
 1487 (<U10DC>, <U10AC>); (<U10DD>, <U10AD>); (<U10DE>, <U10AE>); (<U10DF>, <U10AF>); /  
 1488 (<U10E0>, <U10B0>); (<U10E1>, <U10B1>); (<U10E2>, <U10B2>); (<U10E3>, <U10B3>); /  
 1489 (<U10E4>, <U10B4>); (<U10E5>, <U10B5>); (<U10E6>, <U10B6>); (<U10E7>, <U10B7>); /  
 1490 (<U10E8>, <U10B8>); (<U10E9>, <U10B9>); (<U10EA>, <U10BA>); (<U10EB>, <U10BB>); /  
 1491 (<U10EC>, <U10BC>); (<U10ED>, <U10BD>); (<U10EE>, <U10BE>); (<U10EF>, <U10BF>); /  
 1492 (<U10F0>, <U10C0>); (<U10F1>, <U10C1>); (<U10F2>, <U10C2>); (<U10F3>, <U10C3>); /  
 1493 (<U10F4>, <U10C4>); (<U10F5>, <U10C5>); /  
 1494 (<U1E01>, <U1E00>); (<U1E03>, <U1E02>); (<U1E05>, <U1E04>); /  
 1495 (<U1E07>, <U1E06>); (<U1E09>, <U1E08>); (<U1E0B>, <U1E0A>); (<U1E0D>, <U1E0C>); /  
 1496 (<U1E0F>, <U1E0E>); (<U1E11>, <U1E10>); (<U1E13>, <U1E12>); (<U1E15>, <U1E14>); /  
 1497 (<U1E17>, <U1E16>); (<U1E19>, <U1E18>); (<U1E1B>, <U1E1A>); (<U1E1D>, <U1E1C>); /  
 1498 (<U1E1F>, <U1E1E>); (<U1E21>, <U1E20>); (<U1E23>, <U1E22>); (<U1E25>, <U1E24>); /  
 1499 (<U1E27>, <U1E26>); (<U1E29>, <U1E28>); (<U1E2B>, <U1E2A>); (<U1E2D>, <U1E2C>); /  
 1500 (<U1E2F>, <U1E2E>); (<U1E31>, <U1E30>); (<U1E33>, <U1E32>); (<U1E35>, <U1E34>); /

1501 (<U1E37>, <U1E36>); (<U1E39>, <U1E38>); (<U1E3B>, <U1E3A>); (<U1E3D>, <U1E3C>); /  
 1502 (<U1E3F>, <U1E3E>); (<U1E41>, <U1E40>); (<U1E43>, <U1E42>); (<U1E45>, <U1E44>); /  
 1503 (<U1E47>, <U1E46>); (<U1E49>, <U1E48>); (<U1E4B>, <U1E4A>); (<U1E4D>, <U1E4C>); /  
 1504 (<U1E4F>, <U1E4E>); (<U1E51>, <U1E50>); (<U1E53>, <U1E52>); (<U1E55>, <U1E54>); /  
 1505 (<U1E57>, <U1E56>); (<U1E59>, <U1E58>); (<U1E5B>, <U1E5A>); (<U1E5D>, <U1E5C>); /  
 1506 (<U1E5F>, <U1E5E>); (<U1E61>, <U1E60>); (<U1E63>, <U1E62>); (<U1E65>, <U1E64>); /  
 1507 (<U1E67>, <U1E66>); (<U1E69>, <U1E68>); (<U1E6B>, <U1E6A>); (<U1E6D>, <U1E6C>); /  
 1508 (<U1E6F>, <U1E6E>); (<U1E71>, <U1E70>); (<U1E73>, <U1E72>); (<U1E75>, <U1E74>); /  
 1509 (<U1E77>, <U1E76>); (<U1E79>, <U1E78>); (<U1E7B>, <U1E7A>); (<U1E7D>, <U1E7C>); /  
 1510 (<U1E7F>, <U1E7E>); (<U1E81>, <U1E80>); (<U1E83>, <U1E82>); (<U1E85>, <U1E84>); /  
 1511 (<U1E87>, <U1E86>); (<U1E89>, <U1E88>); (<U1E8B>, <U1E8A>); (<U1E8D>, <U1E8C>); /  
 1512 (<U1E8F>, <U1E8E>); (<U1E91>, <U1E90>); (<U1E93>, <U1E92>); (<U1E95>, <U1E94>); /  
 1513 (<U1E9B>, <U1E9A>); /  
 1514 (<U1EA1>, <U1EA0>); (<U1EA3>, <U1EA2>); (<U1EA5>, <U1EA4>); (<U1EA7>, <U1EA6>); /  
 1515 (<U1EA9>, <U1EA8>); (<U1EAB>, <U1EAA>); (<U1EAD>, <U1EAC>); (<U1EAF>, <U1EAE>); /  
 1516 (<U1EB1>, <U1EB0>); (<U1EB3>, <U1EB2>); (<U1EB5>, <U1EB4>); (<U1EB7>, <U1EB6>); /  
 1517 (<U1EB9>, <U1EB8>); (<U1EBB>, <U1EBA>); (<U1EBD>, <U1EBC>); (<U1EBF>, <U1EBE>); /  
 1518 (<U1EC1>, <U1EC0>); (<U1EC3>, <U1EC2>); (<U1EC5>, <U1EC4>); (<U1EC7>, <U1EC6>); /  
 1519 (<U1EC9>, <U1EC8>); (<U1ECB>, <U1ECA>); (<U1ECD>, <U1ECC>); (<U1EEF>, <U1EEC>); /  
 1520 (<U1ED1>, <U1ED0>); (<U1ED3>, <U1ED2>); (<U1ED5>, <U1ED4>); (<U1ED7>, <U1ED6>); /  
 1521 (<U1ED9>, <U1ED8>); (<U1EDB>, <U1EDA>); (<U1EDD>, <U1EDC>); (<U1EDF>, <U1EDE>); /  
 1522 (<U1EE1>, <U1EE0>); (<U1EE3>, <U1EE2>); (<U1EE5>, <U1EE4>); (<U1EE7>, <U1EE6>); /  
 1523 (<U1EE9>, <U1EE8>); (<U1EEB>, <U1EEA>); (<U1EED>, <U1EEC>); (<U1EEF>, <U1EEE>); /  
 1524 (<U1EF1>, <U1EF0>); (<U1EF3>, <U1EF2>); (<U1EF5>, <U1EF4>); (<U1EF7>, <U1EF6>); /  
 1525 (<U1EF9>, <U1EF8>); (<U1F00>, <U1F08>); (<U1F01>, <U1F09>); (<U1F02>, <U1F0A>); /  
 1526 (<U1F03>, <U1F0B>); (<U1F04>, <U1F0C>); (<U1F05>, <U1F0D>); (<U1F06>, <U1F0E>); /  
 1527 (<U1F07>, <U1F0F>); (<U1F10>, <U1F18>); (<U1F11>, <U1F19>); (<U1F12>, <U1F1A>); /  
 1528 (<U1F13>, <U1F1B>); (<U1F14>, <U1F1C>); (<U1F15>, <U1F1D>); (<U1F20>, <U1F28>); /  
 1529 (<U1F21>, <U1F29>); (<U1F22>, <U1F2A>); (<U1F23>, <U1F2B>); (<U1F24>, <U1F2C>); /  
 1530 (<U1F25>, <U1F2D>); (<U1F26>, <U1F2E>); (<U1F27>, <U1F2F>); (<U1F30>, <U1F38>); /  
 1531 (<U1F31>, <U1F39>); (<U1F32>, <U1F3A>); (<U1F33>, <U1F3B>); (<U1F34>, <U1F3C>); /  
 1532 (<U1F35>, <U1F3D>); (<U1F36>, <U1F3E>); (<U1F37>, <U1F3F>); (<U1F40>, <U1F48>); /  
 1533 (<U1F41>, <U1F49>); (<U1F42>, <U1F4A>); (<U1F43>, <U1F4B>); (<U1F44>, <U1F4C>); /  
 1534 (<U1F45>, <U1F4D>); (<U1F51>, <U1F59>); (<U1F53>, <U1F5B>); (<U1F55>, <U1F5D>); /  
 1535 (<U1F57>, <U1F5F>); (<U1F60>, <U1F68>); (<U1F61>, <U1F69>); (<U1F62>, <U1F6A>); /  
 1536 (<U1F63>, <U1F6B>); (<U1F64>, <U1F6C>); (<U1F65>, <U1F6D>); (<U1F66>, <U1F6E>); /  
 1537 (<U1F67>, <U1F6F>); (<U1F70>, <U1FBA>); (<U1F71>, <U1FBB>); (<U1F72>, <U1FC8>); /  
 1538 (<U1F73>, <U1FC9>); (<U1F74>, <U1FCA>); (<U1F75>, <U1FCB>); (<U1F76>, <U1FDA>); /  
 1539 (<U1F77>, <U1FDB>); (<U1F78>, <U1FF8>); (<U1F79>, <U1FF9>); (<U1F7A>, <U1FEA>); /  
 1540 (<U1F7B>, <U1FEB>); (<U1F7C>, <U1FFA>); (<U1F7D>, <U1FFB>); (<U1F80>, <U1F88>); /  
 1541 (<U1F81>, <U1F89>); (<U1F82>, <U1F8A>); (<U1F83>, <U1F8B>); (<U1F84>, <U1F8C>); /  
 1542 (<U1F85>, <U1F8D>); (<U1F86>, <U1F8E>); (<U1F87>, <U1F8F>); (<U1F90>, <U1F98>); /  
 1543 (<U1F91>, <U1F99>); (<U1F92>, <U1F9A>); (<U1F93>, <U1F9B>); (<U1F94>, <U1F9C>); /  
 1544 (<U1F95>, <U1F9D>); (<U1F96>, <U1F9E>); (<U1F97>, <U1F9F>); (<U1FA0>, <U1FA8>); /  
 1545 (<U1FA1>, <U1FA9>); (<U1FA2>, <U1FAA>); (<U1FA3>, <U1FAB>); (<U1FA4>, <U1FAC>); /  
 1546 (<U1FA5>, <U1FAD>); (<U1FA6>, <U1FAE>); (<U1FA7>, <U1FAF>); (<U1FB0>, <U1FB8>); /  
 1547 (<U1FB1>, <U1FB9>); (<U1FB3>, <U1FBC>); (<U1FC3>, <U1FCC>); (<U1FD0>, <U1FD8>); /  
 1548 (<U1FD1>, <U1FD9>); (<U1FE0>, <U1FE8>); (<U1FE1>, <U1FE9>); (<U1FE5>, <U1FEC>); /  
 1549 (<U1FF3>, <U1FFC>)  
 1550 /  
 1551 tolower /  
 1552 (<U0041>, <U0061>); (<U0042>, <U0062>); (<U0043>, <U0063>); (<U0044>, <U0064>); /  
 1553 (<U0045>, <U0065>); (<U0046>, <U0066>); (<U0047>, <U0067>); (<U0048>, <U0068>); /  
 1554 (<U0049>, <U0069>); (<U004A>, <U006A>); (<U004B>, <U006B>); (<U004C>, <U006C>); /  
 1555 (<U004D>, <U006D>); (<U004E>, <U006E>); (<U004F>, <U006F>); (<U0050>, <U0070>); /  
 1556 (<U0051>, <U0071>); (<U0052>, <U0072>); (<U0053>, <U0073>); (<U0054>, <U0074>); /  
 1557 (<U0055>, <U0075>); (<U0056>, <U0076>); (<U0057>, <U0077>); (<U0058>, <U0078>); /  
 1558 (<U0059>, <U0079>); (<U005A>, <U007A>); (<U00C0>, <U00E0>); (<U00C1>, <U00E1>); /  
 1559 (<U00C2>, <U00E2>); (<U00C3>, <U00E3>); (<U00C4>, <U00E4>); (<U00C5>, <U00E5>); /  
 1560 (<U00C6>, <U00E6>); (<U00C7>, <U00E7>); (<U00C8>, <U00E8>); (<U00C9>, <U00E9>); /  
 1561 (<U00CA>, <U00EA>); (<U00CB>, <U00EB>); (<U00CC>, <U00EC>); (<U00CD>, <U00ED>); /  
 1562 (<U00CE>, <U00EE>); (<U00CF>, <U00EF>); (<U00D0>, <U00F0>); (<U00D1>, <U00F1>); /  
 1563 (<U00D2>, <U00F2>); (<U00D3>, <U00F3>); (<U00D4>, <U00F4>); (<U00D5>, <U00F5>); /  
 1564 (<U00D6>, <U00F6>); (<U00D8>, <U00F8>); (<U00D9>, <U00F9>); (<U00DA>, <U00FA>); /  
 1565 (<U00DB>, <U00FB>); (<U00DC>, <U00FC>); (<U00DD>, <U00FD>); (<U00DE>, <U00FE>); /  
 1566 (<U0178>, <U00FF>); (<U0100>, <U0101>); (<U0102>, <U0103>); (<U0104>, <U0105>); /  
 1567 (<U0106>, <U0107>); (<U0108>, <U0109>); (<U010A>, <U010B>); (<U010C>, <U010D>); /  
 1568 (<U010E>, <U010F>); (<U0110>, <U0111>); (<U0112>, <U0113>); (<U0114>, <U0115>); /  
 1569 (<U0116>, <U0117>); (<U0118>, <U0119>); (<U011A>, <U011B>); (<U011C>, <U011D>); /  
 1570 (<U011E>, <U011F>); (<U0120>, <U0121>); (<U0122>, <U0123>); (<U0124>, <U0125>); /  
 1571 (<U0126>, <U0127>); (<U0128>, <U0129>); (<U012A>, <U012B>); (<U012C>, <U012D>); /  
 1572 (<U012E>, <U012F>); (<U0132>, <U0133>); (<U0134>, <U0135>); (<U0136>, <U0137>); /  
 1573 (<U0139>, <U013A>); (<U013B>, <U013C>); (<U013D>, <U013E>); (<U013F>, <U0140>); /  
 1574 (<U0141>, <U0142>); (<U0143>, <U0144>); (<U0145>, <U0146>); (<U0147>, <U0148>); /  
 1575 (<U014A>, <U014B>); (<U014C>, <U014D>); (<U014E>, <U014F>); (<U0150>, <U0151>); /  
 1576 (<U0152>, <U0153>); (<U0154>, <U0155>); (<U0156>, <U0157>); (<U0158>, <U0159>); /  
 1577 (<U015A>, <U015B>); (<U015C>, <U015D>); (<U015E>, <U015F>); (<U0160>, <U0161>); /  
 1578 (<U0162>, <U0163>); (<U0164>, <U0165>); (<U0166>, <U0167>); (<U0168>, <U0169>); /  
 1579 (<U016A>, <U016B>); (<U016C>, <U016D>); (<U016E>, <U016F>); (<U0170>, <U0171>); /  
 (<U0172>, <U0173>); (<U0174>, <U0175>); (<U0176>, <U0177>); (<U0179>, <U017A>); /

1580 (<U017B>, <U017C>); (<U017D>, <U017E>); (<U0182>, <U0183>); (<U0184>, <U0185>); /  
 1581 (<U0187>, <U0188>); (<U0256>, <U0189>); (<U018B>, <U018C>); (<U018E>, <U01DD>); /  
 1582 (<U0191>, <U0192>); (<U0198>, <U0199>); (<U01A0>, <U01A1>); (<U01A2>, <U01A3>); /  
 1583 (<U01A4>, <U01A5>); (<U01A7>, <U01A8>); (<U01AC>, <U01AD>); (<U01AF>, <U01B0>); /  
 1584 (<U01B3>, <U01B4>); (<U01B5>, <U01B6>); (<U01B8>, <U01B9>); (<U01BC>, <U01BD>); /  
 1585 (<U01C4>, <U01C6>); (<U01C5>, <U01C6>); (<U01C7>, <U01C9>); /  
 1586 (<U01C8>, <U01C9>); (<U01CA>, <U01CC>); (<U01CB>, <U01CC>); /  
 1587 (<U01CD>, <U01CE>); (<U01CF>, <U01D0>); (<U01D1>, <U01D2>); /  
 1588 (<U01D3>, <U01D4>); (<U01D5>, <U01D6>); (<U01D7>, <U01D8>); (<U01D9>, <U01DA>); /  
 1589 (<U01DB>, <U01DC>); (<U01DE>, <U01DF>); (<U01E0>, <U01E1>); (<U01E2>, <U01E3>); /  
 1590 (<U01E4>, <U01E5>); (<U01E6>, <U01E7>); (<U01E8>, <U01E9>); (<U01EA>, <U01EB>); /  
 1591 (<U01EC>, <U01ED>); (<U01EE>, <U01EF>); (<U01F1>, <U01F3>); /  
 1592 (<U01F2>, <U01F3>); (<U01F4>, <U01F5>); (<U01FA>, <U01FB>); (<U01FC>, <U01FD>); /  
 1593 (<U01FE>, <U01FF>); (<U0200>, <U0201>); (<U0202>, <U0203>); (<U0204>, <U0205>); /  
 1594 (<U0206>, <U0207>); (<U0208>, <U0209>); (<U020A>, <U020B>); (<U020C>, <U020D>); /  
 1595 (<U020E>, <U020F>); (<U0210>, <U0211>); (<U0212>, <U0213>); (<U0214>, <U0215>); /  
 1596 (<U0216>, <U0217>); (<U0181>, <U0253>); (<U0186>, <U0254>); (<U018A>, <U0257>); /  
 1597 (<U018E>, <U0258>); (<U018F>, <U0259>); (<U0190>, <U025B>); (<U0193>, <U0260>); /  
 1598 (<U0194>, <U0263>); (<U0197>, <U0268>); (<U0196>, <U0269>); (<U019C>, <U026F>); /  
 1599 (<U019D>, <U0272>); (<U01A9>, <U0283>); (<U01AE>, <U0288>); (<U01B1>, <U028A>); /  
 1600 (<U01B2>, <U028B>); (<U01B7>, <U0292>); (<U0386>, <U03AC>); (<U0388>, <U03AD>); /  
 1601 (<U0389>, <U03AE>); (<U038A>, <U03AF>); (<U0391>, <U03B1>); (<U0392>, <U03B2>); /  
 1602 (<U0393>, <U03B3>); (<U0394>, <U03B4>); (<U0395>, <U03B5>); (<U0396>, <U03B6>); /  
 1603 (<U0397>, <U03B7>); (<U0398>, <U03B8>); (<U0399>, <U03B9>); (<U039A>, <U03BA>); /  
 1604 (<U039B>, <U03BB>); (<U039C>, <U03BC>); (<U039D>, <U03BD>); (<U039E>, <U03BE>); /  
 1605 (<U039F>, <U03BF>); (<U03A0>, <U03C0>); (<U03A1>, <U03C1>); (<U03A3>, <U03C3>); /  
 1606 (<U03A4>, <U03C4>); (<U03A5>, <U03C5>); (<U03A6>, <U03C6>); (<U03A7>, <U03C7>); /  
 1607 (<U03A8>, <U03C8>); (<U03A9>, <U03C9>); (<U03AA>, <U03CA>); (<U03AB>, <U03CB>); /  
 1608 (<U038C>, <U03CC>); (<U038E>, <U03CD>); (<U038F>, <U03CE>); (<U03E2>, <U03E3>); /  
 1609 (<U03E4>, <U03E5>); (<U03E6>, <U03E7>); (<U03E8>, <U03E9>); (<U03EA>, <U03EB>); /  
 1610 (<U03EC>, <U03ED>); (<U03EE>, <U03EF>); (<U0410>, <U0430>); /  
 1611 (<U0411>, <U0431>); (<U0412>, <U0432>); (<U0413>, <U0433>); (<U0414>, <U0434>); /  
 1612 (<U0415>, <U0435>); (<U0416>, <U0436>); (<U0417>, <U0437>); (<U0418>, <U0438>); /  
 1613 (<U0419>, <U0439>); (<U041A>, <U043A>); (<U041B>, <U043B>); (<U041C>, <U043C>); /  
 1614 (<U041D>, <U043D>); (<U041E>, <U043E>); (<U041F>, <U043F>); (<U0420>, <U0440>); /  
 1615 (<U0421>, <U0441>); (<U0422>, <U0442>); (<U0423>, <U0443>); (<U0424>, <U0444>); /  
 1616 (<U0425>, <U0445>); (<U0426>, <U0446>); (<U0427>, <U0447>); (<U0428>, <U0448>); /  
 1617 (<U0429>, <U0449>); (<U042A>, <U044A>); (<U042B>, <U044B>); (<U042C>, <U044C>); /  
 1618 (<U042D>, <U044D>); (<U042E>, <U044E>); (<U042F>, <U044F>); (<U0401>, <U0451>); /  
 1619 (<U0402>, <U0452>); (<U0403>, <U0453>); (<U0404>, <U0454>); (<U0405>, <U0455>); /  
 1620 (<U0406>, <U0456>); (<U0407>, <U0457>); (<U0408>, <U0458>); (<U0409>, <U0459>); /  
 1621 (<U040A>, <U045A>); (<U040B>, <U045B>); (<U040C>, <U045C>); (<U040E>, <U045E>); /  
 1622 (<U040F>, <U045F>); (<U0460>, <U0461>); (<U0462>, <U0463>); (<U0464>, <U0465>); /  
 1623 (<U0466>, <U0467>); (<U0468>, <U0469>); (<U046A>, <U046B>); (<U046C>, <U046D>); /  
 1624 (<U046E>, <U046F>); (<U0470>, <U0471>); (<U0472>, <U0473>); (<U0474>, <U0475>); /  
 1625 (<U0476>, <U0477>); (<U0478>, <U0479>); (<U047A>, <U047B>); (<U047C>, <U047D>); /  
 1626 (<U047E>, <U047F>); (<U0480>, <U0481>); (<U0490>, <U0491>); (<U0492>, <U0493>); /  
 1627 (<U0494>, <U0495>); (<U0496>, <U0497>); (<U0498>, <U0499>); (<U049A>, <U049B>); /  
 1628 (<U049C>, <U049D>); (<U049E>, <U049F>); (<U04A0>, <U04A1>); (<U04A2>, <U04A3>); /  
 1629 (<U04A4>, <U04A5>); (<U04A6>, <U04A7>); (<U04A8>, <U04A9>); (<U04AA>, <U04AB>); /  
 1630 (<U04AC>, <U04AD>); (<U04AE>, <U04AF>); (<U04B0>, <U04B1>); (<U04B2>, <U04B3>); /  
 1631 (<U04B4>, <U04B5>); (<U04B6>, <U04B7>); (<U04B8>, <U04B9>); (<U04BA>, <U04BB>); /  
 1632 (<U04BC>, <U04BD>); (<U04BE>, <U04BF>); (<U04C1>, <U04C2>); (<U04C3>, <U04C4>); /  
 1633 (<U04C7>, <U04C8>); (<U04CB>, <U04CC>); (<U04D0>, <U04D1>); (<U04D2>, <U04D3>); /  
 1634 (<U04D4>, <U04D5>); (<U04D6>, <U04D7>); (<U04D8>, <U04D9>); (<U04DA>, <U04DB>); /  
 1635 (<U04DC>, <U04DD>); (<U04DE>, <U04DF>); (<U04E0>, <U04E1>); (<U04E2>, <U04E3>); /  
 1636 (<U04EA>, <U04E5>); (<U04E6>, <U04E7>); (<U04E8>, <U04E9>); (<U04EA>, <U04EB>); /  
 1637 (<U04EE>, <U04EF>); (<U04F0>, <U04F1>); (<U04F2>, <U04F3>); (<U04F4>, <U04F5>); /  
 1638 (<U04F8>, <U04F9>); (<U0531>, <U0561>); (<U0532>, <U0562>); (<U0533>, <U0563>); /  
 1639 (<U0534>, <U0564>); (<U0535>, <U0565>); (<U0536>, <U0566>); (<U0537>, <U0567>); /  
 1640 (<U0538>, <U0568>); (<U0539>, <U0569>); (<U053A>, <U056A>); (<U053B>, <U056B>); /  
 1641 (<U053C>, <U056C>); (<U053D>, <U056D>); (<U053E>, <U056E>); (<U053F>, <U056F>); /  
 1642 (<U0540>, <U0570>); (<U0541>, <U0571>); (<U0542>, <U0572>); (<U0543>, <U0573>); /  
 1643 (<U0544>, <U0574>); (<U0545>, <U0575>); (<U0546>, <U0576>); (<U0547>, <U0577>); /  
 1644 (<U0548>, <U0578>); (<U0549>, <U0579>); (<U054A>, <U057A>); (<U054B>, <U057B>); /  
 1645 (<U054C>, <U057C>); (<U054D>, <U057D>); (<U054E>, <U057E>); (<U054F>, <U057F>); /  
 1646 (<U0550>, <U0580>); (<U0551>, <U0581>); (<U0552>, <U0582>); (<U0553>, <U0583>); /  
 1647 (<U0554>, <U0584>); (<U0555>, <U0585>); (<U0556>, <U0586>); /  
 1648 (<U10A0>, <U10D0>); (<U10A1>, <U10D1>); (<U10A2>, <U10D2>); (<U10A3>, <U10D3>); /  
 1649 (<U10A4>, <U10D4>); (<U10A5>, <U10D5>); (<U10A6>, <U10D6>); (<U10A7>, <U10D7>); /  
 1650 (<U10A8>, <U10D8>); (<U10A9>, <U10D9>); (<U10AA>, <U10DA>); (<U10AB>, <U10DB>); /  
 1651 (<U10AC>, <U10DC>); (<U10AD>, <U10DD>); (<U10AE>, <U10DE>); (<U10AF>, <U10DF>); /  
 1652 (<U10B0>, <U10E0>); (<U10B1>, <U10E1>); (<U10B2>, <U10E2>); (<U10B3>, <U10E3>); /  
 1653 (<U10B4>, <U10E4>); (<U10B5>, <U10E5>); (<U10B6>, <U10E6>); (<U10B7>, <U10E7>); /  
 1654 (<U10B8>, <U10E8>); (<U10B9>, <U10E9>); (<U10BA>, <U10EA>); (<U10BB>, <U10EB>); /  
 1655 (<U10BC>, <U10EC>); (<U10BD>, <U10ED>); (<U10BE>, <U10EE>); (<U10BF>, <U10EF>); /  
 1656 (<U10C0>, <U10F0>); (<U10C1>, <U10F1>); (<U10C2>, <U10F2>); (<U10C3>, <U10F3>); /

```

1657 (<U10C4>,<U10F4>);(<U10C5>,<U10F5>);/
1658 (<U1E00>,<U1E01>);/
1659 (<U1E02>,<U1E03>);(<U1E04>,<U1E05>);(<U1E06>,<U1E07>);(<U1E08>,<U1E09>);/
1660 (<U1E0A>,<U1E0B>);(<U1E0C>,<U1E0D>);(<U1E0E>,<U1E0F>);(<U1E10>,<U1E11>);/
1661 (<U1E12>,<U1E13>);(<U1E14>,<U1E15>);(<U1E16>,<U1E17>);(<U1E18>,<U1E19>);/
1662 (<U1E1A>,<U1E1B>);(<U1E1C>,<U1E1D>);(<U1E1E>,<U1E1F>);(<U1E20>,<U1E21>);/
1663 (<U1E22>,<U1E23>);(<U1E24>,<U1E25>);(<U1E26>,<U1E27>);(<U1E28>,<U1E29>);/
1664 (<U1E2A>,<U1E2B>);(<U1E2C>,<U1E2D>);(<U1E2E>,<U1E2F>);(<U1E30>,<U1E31>);/
1665 (<U1E32>,<U1E33>);(<U1E34>,<U1E35>);(<U1E36>,<U1E37>);(<U1E38>,<U1E39>);/
1666 (<U1E3A>,<U1E3B>);(<U1E3C>,<U1E3D>);(<U1E3E>,<U1E3F>);(<U1E40>,<U1E41>);/
1667 (<U1E42>,<U1E43>);(<U1E44>,<U1E45>);(<U1E46>,<U1E47>);(<U1E48>,<U1E49>);/
1668 (<U1E4A>,<U1E4B>);(<U1E4C>,<U1E4D>);(<U1E4E>,<U1E4F>);(<U1E50>,<U1E51>);/
1669 (<U1E52>,<U1E53>);(<U1E54>,<U1E55>);(<U1E56>,<U1E57>);(<U1E58>,<U1E59>);/
1670 (<U1E5A>,<U1E5B>);(<U1E5C>,<U1E5D>);(<U1E5E>,<U1E5F>);(<U1E60>,<U1E61>);/
1671 (<U1E62>,<U1E63>);(<U1E64>,<U1E65>);(<U1E66>,<U1E67>);(<U1E68>,<U1E69>);/
1672 (<U1E6A>,<U1E6B>);(<U1E6C>,<U1E6D>);(<U1E6E>,<U1E6F>);(<U1E70>,<U1E71>);/
1673 (<U1E72>,<U1E73>);(<U1E74>,<U1E75>);(<U1E76>,<U1E77>);(<U1E78>,<U1E79>);/
1674 (<U1E7A>,<U1E7B>);(<U1E7C>,<U1E7D>);(<U1E7E>,<U1E7F>);(<U1E80>,<U1E81>);/
1675 (<U1E82>,<U1E83>);(<U1E84>,<U1E85>);(<U1E86>,<U1E87>);(<U1E88>,<U1E89>);/
1676 (<U1E8A>,<U1E8B>);(<U1E8C>,<U1E8D>);(<U1E8E>,<U1E8F>);(<U1E90>,<U1E91>);/
1677 (<U1E92>,<U1E93>);(<U1E94>,<U1E95>);(<U1EA0>,<U1EA1>);(<U1EA2>,<U1EA3>);/
1678 (<U1EA4>,<U1EA5>);(<U1EA6>,<U1EA7>);(<U1EA8>,<U1EA9>);(<U1EAA>,<U1EAB>);/
1679 (<U1EAC>,<U1EAD>);(<U1EAE>,<U1EAF>);(<U1EB0>,<U1EB1>);(<U1EB2>,<U1EB3>);/
1680 (<U1EB4>,<U1EB5>);(<U1EB6>,<U1EB7>);(<U1EB8>,<U1EB9>);(<U1EBA>,<U1EBB>);/
1681 (<U1EBC>,<U1EBD>);(<U1EBE>,<U1EBF>);(<U1EC0>,<U1EC1>);(<U1EC2>,<U1EC3>);/
1682 (<U1EC4>,<U1EC5>);(<U1EC6>,<U1EC7>);(<U1EC8>,<U1EC9>);(<U1ECA>,<U1ECB>);/
1683 (<U1ECC>,<U1ECD>);(<U1ECE>,<U1ECF>);(<U1ED0>,<U1ED1>);(<U1ED2>,<U1ED3>);/
1684 (<U1ED4>,<U1ED5>);(<U1ED6>,<U1ED7>);(<U1ED8>,<U1ED9>);(<U1EDA>,<U1EDB>);/
1685 (<U1EDC>,<U1EDD>);(<U1EDE>,<U1EDF>);(<U1EE0>,<U1EE1>);(<U1EE2>,<U1EE3>);/
1686 (<U1EE4>,<U1EE5>);(<U1EE6>,<U1EE7>);(<U1EE8>,<U1EE9>);(<U1EEA>,<U1EEB>);/
1687 (<U1EEC>,<U1EED>);(<U1EEE>,<U1EEF>);(<U1EF0>,<U1EF1>);(<U1EF2>,<U1EF3>);/
1688 (<U1EF4>,<U1EF5>);(<U1EF6>,<U1EF7>);(<U1EF8>,<U1EF9>);(<U1F08>,<U1F00>);/
1689 (<U1F09>,<U1F01>);(<U1F0A>,<U1F02>);(<U1F0B>,<U1F03>);(<U1F0C>,<U1F04>);/
1690 (<U1F0D>,<U1F05>);(<U1F0E>,<U1F06>);(<U1F0F>,<U1F07>);(<U1F18>,<U1F10>);/
1691 (<U1F19>,<U1F11>);(<U1F1A>,<U1F12>);(<U1F1B>,<U1F13>);(<U1F1C>,<U1F14>);/
1692 (<U1F1D>,<U1F15>);(<U1F28>,<U1F20>);(<U1F29>,<U1F21>);(<U1F2A>,<U1F22>);/
1693 (<U1F2B>,<U1F23>);(<U1F2C>,<U1F24>);(<U1F2D>,<U1F25>);(<U1F2E>,<U1F26>);/
1694 (<U1F2F>,<U1F27>);(<U1F38>,<U1F30>);(<U1F39>,<U1F31>);(<U1F3A>,<U1F32>);/
1695 (<U1F3B>,<U1F33>);(<U1F3C>,<U1F34>);(<U1F3D>,<U1F35>);(<U1F3E>,<U1F36>);/
1696 (<U1F3F>,<U1F37>);(<U1F48>,<U1F40>);(<U1F49>,<U1F41>);(<U1F4A>,<U1F42>);/
1697 (<U1F4B>,<U1F43>);(<U1F4C>,<U1F44>);(<U1F4D>,<U1F45>);(<U1F59>,<U1F51>);/
1698 (<U1F5B>,<U1F53>);(<U1F5D>,<U1F55>);(<U1F5F>,<U1F57>);(<U1F68>,<U1F60>);/
1699 (<U1F69>,<U1F61>);(<U1F6A>,<U1F62>);(<U1F6B>,<U1F63>);(<U1F6C>,<U1F64>);/
1700 (<U1F6D>,<U1F65>);(<U1F6E>,<U1F66>);(<U1F6F>,<U1F67>);(<U1FBA>,<U1F70>);/
1701 (<U1FBB>,<U1F71>);(<U1FC8>,<U1F72>);(<U1FC9>,<U1F73>);(<U1FCA>,<U1F74>);/
1702 (<U1FCB>,<U1F75>);(<U1FDA>,<U1F76>);(<U1FDB>,<U1F77>);(<U1FF8>,<U1F78>);/
1703 (<U1FF9>,<U1F79>);(<U1FEA>,<U1F7A>);(<U1FEB>,<U1F7B>);(<U1FFA>,<U1F7C>);/
1704 (<U1FFB>,<U1F7D>);(<U1F88>,<U1F80>);(<U1F89>,<U1F81>);(<U1F8A>,<U1F82>);/
1705 (<U1F8B>,<U1F83>);(<U1F8C>,<U1F84>);(<U1F8D>,<U1F85>);(<U1F8E>,<U1F86>);/
1706 (<U1F8F>,<U1F87>);(<U1F98>,<U1F90>);(<U1F99>,<U1F91>);(<U1F9A>,<U1F92>);/
1707 (<U1F9B>,<U1F93>);(<U1F9C>,<U1F94>);(<U1F9D>,<U1F95>);(<U1F9E>,<U1F96>);/
1708 (<U1F9F>,<U1F97>);(<U1FA8>,<U1FA0>);(<U1FA9>,<U1FA1>);(<U1FAA>,<U1FA2>);/
1709 (<U1FAB>,<U1FA3>);(<U1FAC>,<U1FA4>);(<U1FAD>,<U1FA5>);(<U1FAE>,<U1FA6>);/
1710 (<U1FAF>,<U1FA7>);(<U1FB8>,<U1FB0>);(<U1FB9>,<U1FB1>);(<U1FBC>,<U1FB3>);/
1711 (<U1FCC>,<U1FC3>);(<U1FD8>,<U1FD0>);(<U1FD9>,<U1FD1>);(<U1FE8>,<U1FE0>);/
1712 (<U1FE9>,<U1FE1>);(<U1FEC>,<U1FE5>);(<U1FFC>,<U1FF3>)
1713
1714 % The "combining" class reflects ISO/IEC 10646-1 annex B.1
1715 % That is, all combining characters (level 2+3).
1716 class "combining"; /
1717 <U0300>..<U036F>; <U20D0>..<U20FF>; <UFE20>..<UFE2F>;/
1718 <U0483>..<U0486>;<U0591>..<U05A1>;<U05A3>..<U05B9>;/
1719 <U05BB>..<U05BD>;<U05BF>;<U05C1>;<U05C2>;<U05C4>;<U064B>..<U0652>;<U0670>;/
1720 <U06D7>..<U06E4>;<U06E7>;<U06E8>;<U06EA>..<U06ED>;<U0901>..<U0903>;<U093C>;/
1721 <U093E>..<U094D>;<U0951>..<U0954>;<U0962>;<U0963>;<U0981>..<U0983>;<U09BC>;/
1722 <U09BE>..<U09C4>;<U09C7>;<U09C8>;<U09CB>..<U09CD>;<U09D7>;<U09E2>;<U09E3>;/
1723 <U0A02>;<U0A3C>;<U0A3E>..<U0A42>;<U0A47>;<U0A48>;<U0A4B>..<U0A4D>;/
1724 <U0A70>;<U0A71>;<U0A81>..<U0A83>;<U0ABC>;<U0ABE>..<U0AC5>;<U0AC7>..<U0AC9>;/
1725 <U0ACB>..<U0ACD>;<U0B01>..<U0B03>;<U0B3C>;<U0B3E>..<U0B43>;<U0B47>;<U0B48>;/
1726 <U0B4B>..<U0B4D>;<U0B56>;<U0B57>;<U0B82>;<U0B83>;<U0BBE>..<U0BC2>;/
1727 <U0BC6>..<U0BC8>;<U0BCA>..<U0BCD>;<U0BD7>;<U0C01>..<U0C03>;<U0C3E>..<U0C44>;/
1728 <U0C46>..<U0C48>;<U0C4A>..<U0C4D>;<U0C55>;<U0C56>;<U0C82>;<U0C83>;/
1729 <U0CBE>..<U0CC4>;<U0CC6>..<U0CC8>;<U0CCA>..<U0CCD>;<U0CD5>;<U0CD6>;/
1730 <U0D02>;<U0D03>;<U0D3E>..<U0D43>;<U0D46>..<U0D48>;<U0D4A>..<U0D4D>;<U0D57>;/
1731 <U0E31>;<U0E34>..<U0E3A>;<U0E47>..<U0E4E>;<U0EB1>;<U0EB4>..<U0EB9>;/
1732 <U0EBB>;<U0EBC>;<U0EC8>..<U0ECD>;<U0F18>;<U0F19>;<U0F35>;<U0F37>;<U0F39>;/
1733 <U0F3E>;<U0F3F>;<U0F71>..<U0F84>;<U0F86>..<U0F89>;<U0F8B>;<U0F90>..<U0F95>;/
1734 <U0F97>;<U0F99>..<U0FAD>;<U0FB1>..<U0FB7>;<U0FB9>;<U302A>..<U302F>;/
1735 <U3099>;<U309A>;<UFB1E>

```

```

1736 %
1737 % The "combining_level3" class reflects ISO/IEC 10646-1 annex B.2
1738 % That is, combining characters of level 3.
1739 class      "combining_level3";      /
1740 <U0300>..

```

### 4.3 LC\_COLLATE

A collation sequence definition defines the relative order between collating elements (characters and multicharacter collating elements) in the FDCC-set. This order is expressed in terms of collation values; i.e., by assigning each element one or more collation values (also known as collation weights). This does not imply that applications shall assign such values, but that ordering of strings using the resultant collation definition in the FDCC-set shall behave as if such assignment is done and used in the collation process. The collation sequence definition is used by regular expressions, pattern matching, and sorting. The following capabilities are provided:

- (1) Multicharacter collating elements. Specification of multicharacter collating elements (i.e., sequences of two or more characters to be collated as an entity).
- (2) User-defined ordering of collating elements. Each collating element shall be assigned a collation value defining its order in the character (or basic) collation sequence. This ordering is used by regular expressions and pattern matching and, unless collation weights are explicitly specified, also as the collation weight to be used in sorting.
- (3) Multiple weights and equivalence classes. Collating elements can be assigned one or more (up to the limit (COLL\_WEIGHTS\_MAX)) collating weights for use in sorting. The first weight is hereafter referred to as the primary weight.
- (4) One-to Many mapping. A single character is mapped into a string of collating elements.
- (5) Many-to-Many substitution. A string of one or more characters is substituted by another string (or an empty string, i.e., the character or characters shall be ignored for collation purposes).
- (6) Equivalence class definition. Two or more collating elements have the same collation value (primary weight).
- (7) Ordering by weights. When two strings are compared to determine their relative order, the two strings are first broken up into a series of collating elements, and each successive pair of elements are compared according to the relative primary weights for the elements. If equal, and more than one weight has been assigned, then the pairs of collating elements are recompared according to the relative subsequent weights, until either a pair of collating elements compare unequal or the weights are exhausted.
- (8) Per section ordering rules. Some cultures order some scripts in a different direction than other scripts, for example in French cultures the Latin script is ordered backwards on the level handling accents, while the Cyrillic script may be ordered forwards. Collections of such scripts or other collections of characters can be handled together in a section, with a specific set of rules applied per section.



- 1790 (9) Easy reordering of characters. ISO/IEC 14651 has a template for collation  
 1791 specification that with just a few modifications can be culturally correct for a  
 1792 specific culture. Here the "reorder-after" keyword gives a convenient way to  
 1793 modify a FDCC-set template.
- 1794 (10) Easy reordering of sections. The template in ISO/IEC 14651 gives an ordering of  
 1795 the sections that may not be culturally acceptable in certain cultures. The keyword  
 1796 "reorder-section-after" gives a convenient way to modify the order of sections in a  
 1797 FDCC-set template.

1798  
 1799 The following keywords shall be defined in a collation sequence definition. Some of them  
 1800 are described in detail in the following subclauses.

1801		
1802	<b>copy</b>	Specify the name of an existing FDCC-set to be used as the source for the definition of this category. If this keyword is specified, only the "reorder-after", "reorder-end", "reorder-sections-after" and "reorder-sections-end" keywords may also be specified. The FDCC-set shall be copied in source form.
1803		
1804		
1805		
1806		
1807		
1808	<b>coll_weight_max</b>	Define as a decimal number the number of collation levels that an interpreting system needs to support for this FDCC-set, this value is elsewhere referred as the COLL_WEIGHT_MAX limit. An interpreting system shall cater for up to 7 collating levels.
1809		
1810		
1811		
1812		
1813	<b>section-symbol</b>	Define a section symbol representing a set of collation order statements. The section is defined with the "order_start" keyword until the next "order_start" or "order_end" keyword. This keyword is optional.
1814		
1815		
1816		
1817		
1818	<b>collating-element</b>	Define a collating-element symbol representing a multicharacter collating element. This keyword is optional.
1819		
1820		
1821	<b>collating-symbol</b>	Define one or more collating symbols for use in collation order statements. This keyword is optional.
1822		
1823	<b>symbol-equivalence</b>	Define a collating-symbol to be equivalent to another defined collating-symbol.
1824		
1825	<b>order_start</b>	Define collation rules. This statement is followed by one or more collation order statements, assigning character collation values and collation weights to collating elements.
1826		
1827		
1828		
1829	<b>order_end</b>	Specify the end of the collation-order statements.
1830	<b>reorder-after</b>	Redefine collating rules. Specify after which collating element the redefinition of collation order shall take order. This statement is followed by one or more collation order statements, reassigning character collation values and collation weights to collating elements.
1831		
1832		
1833		
1834		
1835		
1836	<b>reorder-end</b>	Specify the end of the "reorder-after" collating order statements.
1837		
1838	<b>reorder-section-after</b>	Redefine the order of sections. This statement is followed by one or more section symbols,
1839		

1840		reassigning character collation values and collation
1841		weights to collating elements.
1842	<b>reorder-section-end</b>	Specify the end of the "reorder-sections" section
1843		order statements.
1844		
1845	<b>Toggling keywords:</b>	
1846		
1847	<b>define</b>	defines a toggle.
1848	<b>undef</b>	undefines a toggle.
1849	<b>ifdef</b>	tests a toggle, and if defined uses the following
1850		statements.
1851	<b>ifndef</b>	tests a toggle, and if undefined uses the following
1852		statements.
1853	<b>else</b>	uses the following statements if no preceding
1854		toggling statements have been used.
1855	<b>elif</b>	tests a toggle, and uses the following statements if no
1856		preceding toggling statements have been used, and
1857		the toggle is defined.
1858	<b>endif</b>	terminates set of toggling statements.
1859		

#### 1860 4.3.1 Collation statements

1861 The "order\_start" and "replace-after" keywords shall be followed by collating statements.  
 1862 The syntax for the collating statements is

```
1863 "%s %s;%s;...;%s\n",<collating-identifier>,<weight>,<weight>,...
```

1864  
 1865 Each <collating-identifier> shall consist of either a character (in any of the forms defined  
 1866 in 4.0.1), a <collating-element>, a <collating-symbol>, an ellipsis, or the special symbol  
 1867 "UNDEFINED". The order in which collating elements are specified determines the  
 1868 character collation sequence, such that each collating element shall compare less than the  
 1869 elements following it. The NUL character shall compare lower than any other character.

1870  
 1871 A <collating-element> shall be used to specify multicharacter collating elements, and  
 1872 indicates that the character sequence specified via the <collating-element> is to be collated  
 1873 as a unit and in the relative order specified by its place.

1874  
 1875 A <collating-symbol> shall be used to define a position in the relative order for use in  
 1876 weights.

1877  
 1878 The absolute ellipsis symbol ("...") specifies that a sequence of characters shall collate  
 1879 according to their encoded character values. It shall be interpreted as indicating that all  
 1880 characters with a coded character set value higher than the value of the character in the  
 1881 preceding line, and lower than the coded character set value for the character in the  
 1882 following line, in the current coded character set, shall be placed in the character collation  
 1883 order between the previous and the following character in ascending order according to  
 1884 their coded character set values. An initial ellipsis shall be interpreted as if the preceding  
 1885 line specified the <NUL> character, and a trailing ellipsis as if the following line specified  
 1886 the highest coded character set value in the current coded character set. An ellipsis shall  
 1887 be treated as invalid if the preceding or following lines do not specify characters in the  
 1888  
 1889

1890 current coded character set. The use of the ellipsis symbol ties the definition to a specific  
1891 coded character set and may preclude the definition from being portable between  
1892 applications, and is depreciated. Symbolic ellipses may be used as the ellipses symbol, but  
1893 generating symbolic character names, and thus have a better chance of portability between  
1894 applications.

1895  
1896 The symbolic ellipses (".." or "...") specifies a sequence of collating statements. It shall  
1897 be interpreted as indicating that all characters with symbolic names higher than the  
1898 symbolic name of the character in the preceding line, and lower than the coded character  
1899 set value for the character in the following line, shall be placed in the character collation  
1900 order between the previous and the following character in ascending order.

1901  
1902 The symbol "UNDEFINED" shall be interpreted as including all coded character set values  
1903 not specified explicitly or via the ellipsis or one of the symbolic ellipses symbols. Such  
1904 characters shall be inserted in the character collation order at the point indicated by the  
1905 symbol, and in ascending order according to their coded character set values. If no  
1906 "UNDEFINED" symbol is specified, and the current coded character set contains  
1907 characters not specified in this clause, the utility shall issue a warning message and place  
1908 such characters at the end of the character collation order.

1909  
1910 The optional operands for each collation-element shall be used to define the primary,  
1911 secondary, or subsequent weights for the collating element. The first operand specifies the  
1912 relative primary weight, the second the relative secondary weight, and so on. Two or more  
1913 collation-elements can be assigned the same weight; they belong to the same equivalence  
1914 class if they have the same primary weight. Collation shall behave as if, for each weight  
1915 level, "IGNORE"d elements are removed. Then each successive pair of elements shall be  
1916 compared according to the relative weights for the elements. If the two strings compare  
1917 equal, the process shall be repeated for the next weight level, up to the limit  
1918 "COLL\_WEIGHTS\_MAX" of the associated FDCC-set.

1919  
1920 Weights shall be expressed as characters (in any of the forms specified here), <collating-  
1921 symbol>s, <collating-element>s, an ellipsis, or the special symbol "IGNORE". A single  
1922 character, a <collating-symbol>, or a <collating-element> shall represent the relative order  
1923 in the character collating sequence of the character or symbol, rather than the character or  
1924 characters themselves.

1925  
1926 One-to-many mapping is indicated by specifying two or more concatenated characters or  
1927 symbolic names. Thus, if the character <ss> is given the string <s><s> as a weight,  
1928 comparisons shall be performed as if all occurrences of the character <ss> are replaced by  
1929 <s><s>. If it is desirable to define <ss> and <s><s> as an equivalence class, then a  
1930 collating-element must be defined for the string "ss", as in the example below.

1931  
1932 All characters specified via an ellipsis shall by default be assigned unique weights, equal  
1933 to the relative order of characters. Characters specified via an explicit or implicit  
1934 "UNDEFINED" special symbol shall by default be assigned the same primary weight (i.e.,  
1935 belong to the same equivalence class). An ellipsis symbol as a weight shall be interpreted  
1936 to mean that each character in the sequence shall have unique weights, equal to the  
1937 relative order of their character in the character collation sequence. Secondary and  
1938 subsequent weights have unique values. The use of the ellipsis as a weight shall be treated  
1939 as an error if the collating element is neither an ellipsis nor the special symbol

1940 "UNDEFINED".

1941

1942 The special keyword "IGNORE" as a weight shall indicate that when strings are compared  
 1943 using the weights at the level where "IGNORE" is specified, the collating element shall be  
 1944 ignored; i.e., as if the string did not contain the collating element. In regular expressions  
 1945 and pattern matching, all characters that are "IGNORE"d in their primary weight form an  
 1946 equivalence class.

1947

1948 A <comment character> occurring where the delimiter ";" may occur, terminates the  
 1949 collating statement.

1950

1951 An empty operand shall be interpreted as the collating-element itself.

1952

1953 For example, the collation statement

1954

```
<a> <a>;<a>
```

1956

1957 is equal to

1958

```
<a>
```

1959

1960

1961 An ellipsis (absolute or symbolic) can be used as an operand if the collating-element was  
 1962 an ellipsis, and shall be interpreted as the value of each character defined by the ellipsis.

1963

1964

Example:

1965

1966

```
collating-element <ch> from "<c><h>"
collating-element <Ch> from "<C><H>"
order_start      forward;backward
UNDEFINED       IGNORE;IGNORE
<LOW>
<space>         <LOW>;<space>
...
<a>             <a>;<a>
<a'>           <a>;<a'>
<A>            <a>;<A>
<A'>          <a>;<A'>
<ch>           <ch>;<ch>
<Ch>          <ch>;<Ch>
<s>            <s>;<s>
<ss>          "<s><s>"; "<SS><SS>"
order_end
```

1981

1982

1983 This example is interpreted as follows:

1984

- 1985 (1) The UNDEFINED means that all characters not specified in this definition (explicitly or via the  
 1986 ellipsis) shall be ignored.
- 1987 (2) <LOW> defines the first collating weight, and thus the lowest weight in this example.
- 1988 (3) All characters between <space> and <a> shall have the same primary equivalence class <LOW> and  
 1989 individual secondary weights based on their ordinal encoded values. (The use of absolute ellipses is  
 1990 deprecated, but used here to illustrate generic use of ellipses. Symbolic ellipses should be used  
 1991 instead).
- 1992 (4) All characters based on the upper or lowercase character "a" belong to the same primary equivalence  
 1993 class.
- 1994 (5) The multicharacter collating element <c><h> is represented by the collating symbol <ch> and belongs  
 1995 to the same primary equivalence class as the multicharacter collating element <C><H>.
- 1996 (6) The <ss> collating element has two weights on the primary level, and it is in the same primary  
 1997 equivalence class as two consecutive <s>-es; on the secondary level the collating element has two  
 1998 weights of the equivalence class <ss>.

1999

2000

#### 4.3.2 "copy" keyword

2001

2002 This keyword specifies the name of an existing FDCC-set to be used as the source for the  
 2003 definition of this category. The syntax is

2004  
 2005 "copy %s\n", <FDCC-set-name>  
 2006

2007 The <FDCC-set-name> shall consist of one or more characters (in any of the forms  
 2008 defined in 4.0.1). If this keyword is specified, only the "reorder-after", "reorder-end",  
 2009 "reorder-sections-after" and "reorder-sections-end" keywords may also be specified. The  
 2010 FDCC-set shall be copied in source form.

### 2012 4.3.3 "col\_weight\_max" keyword

2014 This keyword defines as a decimal number the number of collation levels that an  
 2015 interpreting system needs to support, this value is elsewhere referred as the  
 2016 COLL\_WEIGHT\_MAX limit. The minimum value is 7. The syntax is

2017  
 2018 "col\_weight\_max %d\n", <value>  
 2019

### 2020 4.3.4 "section-symbol" keyword

2022 This keyword shall be used to define symbols for use in section related statements; such  
 2023 as the "order\_start", and "reorder-sections-after" keywords and section-reordering  
 2024 statements. The syntax is

2025  
 2026 "section-symbol %s\n", <section-symbol>  
 2027

2028 The <section-symbol> shall be a symbolic name, enclosed between angle brackets (< and  
 2029 >), and shall not duplicate any symbolic name in the current charmap (if any), or any  
 2030 other symbolic name defined in this collation definition. A <section-symbol> defined via  
 2031 this keyword is only defined with the LC\_COLLATE category.

2032 Example:

2033 section-symbol <LATIN>  
 2034 section-symbol <ARABIC>  
 2035

### 2037 4.3.5 "collating-element" keyword

2038 In addition to the collating elements in the character set, the collating-element keyword  
 2039 shall be used to define multicharacter collating elements. The syntax is

2040  
 2041 "collating-element %s from %s\n", <collating-symbol>, <string>  
 2042

2043 The <collating-symbol> operand shall be a symbolic name, enclosed between angle  
 2044 brackets (< and >), and shall not duplicate any symbolic name in the current charmap or  
 2045 repertoiremap file (if any), or any other symbolic name defined in this collation definition.  
 2046 The string operand shall be a string of two or more characters that shall collate as an  
 2047 entity. A <collating-element> defined via this keyword is only defined within the  
 2048 LC\_COLLATE category.

2049 Example with ISO/IEC 10646:

2050 collating-element <ch> from "<c><h>"  
 2051 collating-element <e-acute> from "<e><combining-acute>"  
 2052  
 2053

2054 collating-element <aa> from "<a><a>"

2055

2056 Note: The problem of comparing a fully composed character of ISO/IEC 10646 with a  
2057 decomposed representation of the same text is normally handled by the two strings  
2058 comparing equal up to level 3 (the case level) of ISO/IEC 14651, but distinguishing the  
2059 two at the 4th level.

2060

#### 2061 **4.3.6 "collating-symbol" keyword**

2062

2063 This keyword shall be used to define symbols for use in collation sequence statements;  
2064 e.g., between the order\_start and the order\_end keywords. The syntax is

2065

2066 "collating-symbol %s;%s;...%s\n", <collating-symbol>, <collating-symbol> ...

2067

2068 The <collating-symbol> shall be a symbolic name, enclosed between angle brackets (< and  
2069 >), and shall not duplicate any symbolic name in the current charmap (if any), or any  
2070 other symbolic name defined in this collation definition. A <collating-symbol> defined via  
2071 this keyword is only defined with the LC\_COLLATE category. More than one <collating-  
2072 symbol> may be defined with one "collating-symbol" keyword, and symbolic ellipses may  
2073 be used.

2074

2075 Example:

2076 collating-symbol <CAPITAL>

2077 collating-symbol <HIGH>

2078

#### 2079 **4.3.7 "symbol-equivalence" keyword**

2080

2081 This keyword shall be used to define symbols for use in collation sequence statements;  
2082 and assign the same weight as another defined symbol. The syntax is

2083

2084 "symbol-equivalence %s %s\n", <collating-symbol-1>, <collating-symbol-2>

2085

2086 The <collating-symbol-1> and <collating-symbol-2> shall be symbolic names, enclosed  
2087 between angle brackets (< and >). <collating-symbol-1> shall not duplicate any symbolic  
2088 name in the current charmap (if any), or any other symbolic name defined in this collation  
2089 definition. <collating-symbol-2> is defined elsewhere in the LC\_COLLATE category as a  
2090 collating-symbol. The use of <collating-symbol-2> shall be equivalent to using the  
2091 <collating-symbol-2> in the LC\_COLLATE category. A <collating-symbol-1> defined via  
2092 this keyword is only defined with the LC\_COLLATE category.

2093

2094 Example

2095 collating-symbol <CAP>

2096 symbol-equivalence <CAPITAL> <CAP>

2097

#### 2098 **4.3.8 "order\_start" keyword**

2099

2100 The "order\_start" keyword shall precede collation order entries and also defines the  
2101 number of weights for this collation sequence definition, the collation section name and  
2102 other collation rules.

2103

2104 The syntax of the "order\_start" keyword has two forms:

2105

"order\_start %s;%s;...;%s\n", <sort-rules>, <sort-rules> ...

2106

and

2107

"order\_start %s;%s;...;%s\n", <section-symbol>, <sort-rules>, <sort-rules> ...

2108

2109

2110

The operands to the order\_start keyword are optional. If present, the operands define rules to be applied when strings are compared. The first operand may be a <section-symbol> surrounded by "<" and ">" and the set of collating statements following the "order\_start" keyword until the "order\_end" keyword are identified with this <section\_symbol> or another "order\_start" keyword is encountered. The remaining number of operands define how many weights each element is assigned; if no operands are present, one forward operand is assumed. If present, the first operand defines rules to be applied when comparing strings using the first (primary) weight; the second when comparing strings using the second weight, and so on. Operands shall be separated by semicolons (;). Each operand shall consist of one or more collation directives, separated by commas (.). If the number of operands exceeds the (COLL\_WEIGHTS\_MAX) limit, a utility parsing the FDCC-set description shall issue a warning message. The following directives shall be supported:

2111

2112

2113

2114

2115

2116

2117

2118

2119

2120

2121

2122

2123

**forward** Specifies that the direction of scanning a part of a string at a given point in a string is done towards the logical end of the whole string for this weight level.

2124

2125

**backward** Specifies that the direction of scanning a part of a string at a given point in a string is done towards the logical beginning of the whole string for this weight level.

2126

2127

2128

2129

**position** Specifies that comparison operations for the weight level will consider the relative position of non-"IGNORE"d elements in the strings. The string containing a non-"IGNORE"d element after the fewest IGNOREd collating elements from the start of the compare shall collate first. If both strings contain a non-"IGNORE"d character in the same relative position, the collating values assigned to the elements shall determine the ordering. In case of equality, subsequent non-IGNOREd characters shall be considered in the same manner.

2130

2131

2132

2133

2134

2135

2136

2137

The directives "forward" and "backward", and "backward" and "position", are mutually exclusive at a given level.

2138

2139

2140

Examples:

2141

order\_start forward;backward

2142

order\_start <CYRILLIC>;forward;forward

2143

2144

If no operands are specified, a single forward operand shall be assumed.

2145

2146

2147

#### 4.3.9 "order\_end" keyword

2148

2149

The collating order entries shall be terminated with an order\_end keyword.

2150

2151

#### 4.3.10 "reorder-after" keyword

2152

2153

The "reorder-after" keyword shall be used to specify a modification to a copied collation

2154

2155 specification of an existing FDCC-set. There can be more than one "reorder-after"  
 2156 statement in a collating specification. The syntax shall be:

2157  
 2158 "reorder-after %s\n",<collating-symbol>  
 2159

2160 The <collating-symbol> operand shall be a symbolic name, enclosed between angle  
 2161 brackets, and shall be present in the source FDCC-set copied via the "copy" keyword.  
 2162 The "reorder-after" statement is followed by one or more collation statements as described  
 2163 in the "Collating Order" clause (4.3.5), with the exception that the ellipsis symbol (...)  
 2164 shall not be used.

2165  
 2166 Each collation statement reassigns character collation values and collation weights to  
 2167 collating elements existing in the copied collation specification, by removing the collating  
 2168 statement from the copied specification, and inserting the collating element in the collating  
 2169 sequence with the new collation weights after the preceding collating element of the  
 2170 "reorder-after" specification, the first collating element in the collation sequence being the  
 2171 <collating-symbol> specified on the "reorder-after" statement.  
 2172

2173 A "reorder-after" specification is terminated by another "reorder-after" specification or the  
 2174 "reorder-end" statement.

2175

#### 2176 4.3.10.1 Example of "reorder-after"

2177

```

2178 reorder-after <y8>
2179 <U:>      <Y>;<U:>;<CAPITAL>
2180 <u:>      <Y>;<U:>;<SMALL>
2181 reorder-after <z8>
2182 <AE>      <AE>;<NONE>;<CAPITAL>
2183 <ae>      <AE>;<NONE>;<SMALL>
2184 <A:>      <AE>;<DIAERESIS>;<CAPITAL>
2185 <a:>      <AE>;<DIAERESIS>;<SMALL>
2186 <O/>      <O/>;<NONE>;<CAPITAL>
2187 <o/>      <O/>;<NONE>;<SMALL>
2188 <AA>      <AA>;<NONE>;<CAPITAL>
2189 <aa>      <AA>;<NONE>;<SMALL>
2190 reorder-end
  
```

2191

2192 The example is interpreted as follows (using the "i18nrep" repertoire map):

2193

- 2194 1. The collating element <U:> is removed from the copied collating sequence and inserted after <y8> in the  
 2195 collating sequence with the new weights. The collating element <u:> is removed from the copied collating  
 2196 sequence and inserted in the resulting collation sequence after <U:> with the new weights. <y8> is used to  
 2197 indicate the last entry of the <y> letters.  
 2198
- 2199 2. The second "reorder-after" statement terminates the first list of reordering collation identifier entries, and  
 2200 initiates a second list, rearranging the order and weights for the <AE>, <ae>, <A:>, <a:>, <O/>, and <o/>  
 2201 collating elements after the <z8> collating symbol in the copied specification. <z8> is used to indicate the  
 2202 last entry of the <z> letters.  
 2203

2204

- 2205 3. The "reorder-end" statement terminates the second list of reordering entries.

2206

- 2207 4. Thus for the original sequence

2208

2209 ... ( U u Ü ü ) V v W w X x Y y Z z

2210

2210 this example reordering gives



2211 ... U u V v W w X x ( Y y Ü ü ) Z z ( Æ æ Ä ä ) Ø ø Å å

2212

2213 where the parenthesis indicate ordering with the same weight on the first level for multiple upper/lowercase  
2214 pairs.

2215

### 2216 4.3.11 "reorder-end" keyword

2217

2218 The "reorder-end" keyword shall specify the end of a list of collating statements, initiated  
2219 by the "reorder-after" keyword.

2220

### 2221 4.3.12 "reorder-sections-after" keyword

2222

2223 The "reorder-sections-after" keyword shall be used to specify a modification to a copied  
2224 collation specification of an existing FDCC-set. The "reorder-sections-after" statement is  
2225 followed by one or more statements consisting of section reordering statements.

2226

#### 2227 4.3.12.1 section reordering statements

2228

2229 The section reordering statements rearranges the set of collating entries and changes  
2230 sorting rules for the set of collating entries identified by a section symbol in a preceding  
2231 "order\_start" statement. Each section reorder statement has the syntax:

2232

2233 "%s %s;...%s\n", <section-symbol>, <sort-rules>, <sort-rules> ...

2234

2235 The <section-symbol> identifies the set of collating entries, and shall be defined via a  
2236 "section-symbol" keyword.

2237

2238 The <sort-rules> are as described for the "order\_start" keyword. Specified <sort-rules>  
2239 replace the specification for the ordering of the section given on the "order\_start"  
2240 statement identified by the <section-symbol>. The <sort-rules> are optional and <sort-  
2241 rules> not to be changed may be given by empty specifications.

2242

2243 The order of the section reordering statements rearranges the assignment of collation  
2244 entries for the sets of collation entries identified by the <section-symbols> to the order  
2245 that the <section-symbols> occur after the "reorder-sections-after" statement.

2246

2247 The section reordering statements are terminated by a "reorder-sections-end" statement.

2248

#### 2249 4.3.12.2 Example of section reordering

2250

```
2251 copy "i18n"
2252 reorder-sections-after <DIGITS>
2253 <ARABIC>
2254 <LATIN> forward;backward;forward;forward,position
2255 reorder-sections-end
2256
```

2257 This example is interpreted as follows: The LC\_COLLATE category of the "i18n" FDCC-set is copied. Then a  
2258 reordering of all collating statements for the sections <ARABIC> and <LATIN> is done, leaving the rest of the  
2259 sections as they were in the "i18n" FDCC-set. The <ARABIC> section is placed immediately after the <DIGITS>  
2260 section, and the <LATIN> section immediately following the <ARABIC> section. The ordering rules are kept as  
2261 they were in the "i18n" FDCC-set, while the <LATIN> section gets new ordering rules as indicated. The  
2262 "reorder-sections-end" keyword terminates the section reordering statements.

2263

### 2264 4.3.13 "reorder-sections-end" keyword

2265

2266 The "reorder-sections-end" keyword shall specify the end of a list of section symbols,  
2267 initiated by the "reorder-sections-after" keyword.

2268

#### 2269 **4.3.14 Toggling keyword statements**

2270

2271 The toggling keywords "define" and "undef" shall set, respectively unset a toggle. Toggles  
2272 that are not defined, are regarded as unset. The toggle is a string of characters, in any  
2273 form as described in clause 4.0.1. The keywords "ifdef", "ifndef", "elif", "else", and  
2274 "endif" controls the inclusion of LC\_COLLATE keywords and statements, as described in  
2275 the following, and they work in a nesting manner. The toggling keywords are modelled  
2276 after the precompiler in the C standard.

2277

##### 2278 **4.3.14.1 "define" keyword**

2279

2280 This keyword shall be used to set a toggle, for use with other toggling keywords. The  
2281 same toggle may occur with more "define" statements. The syntax is

2282

2283 "define %s\n", <toggle>

2284

##### 2285 **4.3.14.2 "undef" keyword**

2286

2287 This keyword shall be used to unset a toggle, for use with other toggling keywords. The  
2288 same toggle may occur with more "undef" statements. The syntax is

2289

2290 "undef %s\n", <toggle>

2291

##### 2292 **4.3.14.3 "ifdef" keyword**

2293

2294 This keyword shall be used to control the inclusion of the following LC\_COLLATE  
2295 statements, up to a corresponding "elif", "else" or "endif" keyword. If the toggle is set, the  
2296 statements are used, otherwise they are ignored. The syntax is

2297

2298 "ifdef %s\n", <toggle>

2299

##### 2300 **4.3.14.4 "ifndef" keyword**

2301

2302 This keyword shall be used to control the inclusion of the following LC\_COLLATE  
2303 statements, up to a corresponding "elif", "else" or "endif" keyword. If the toggle is unset,  
2304 the statements are used, otherwise they are ignored. The syntax is

2305

2306 "ifndef %s\n", <toggle>

2307

##### 2308 **4.3.14.5 "elif" keyword**

2309

2310 This keyword shall be used to control the inclusion of the following LC\_COLLATE  
2311 statements, up to a corresponding "elif", "else" or "endif" keyword. The keyword shall be  
2312 preceded by a corresponding "ifdef", "ifndef", or "elif" statement and the statement that  
2313 these keyword statements control. If no preceding "ifdef", "ifndef" or "elif" statement has  
2314 been used, and if the toggle is set, the statements are used, otherwise they are ignored.

2315

The syntax is

2316 "elif %s\n", <toggle>  
 2317

#### 2318 4.3.14.6 "else" keyword

2319  
 2320 This keyword shall be used to control the inclusion of the following LC\_COLLATE  
 2321 statements, up to a corresponding "endif" keyword. The keyword shall be preceded by a  
 2322 corresponding "ifdef", "ifndef", or "elif" statement and the statement that these keyword  
 2323 statements control. If no preceding "ifdef", "ifndef" or "elif" statement has been used, the  
 2324 statements are used, otherwise they are ignored. The syntax is

2325  
 2326 "else\n"  
 2327

#### 2328 4.3.14.7 "endif" keyword

2329  
 2330 This keyword shall be used to terminate the control of the inclusion of the preceding  
 2331 LC\_COLLATE statements. The keyword shall be preceded by a corresponding "ifdef",  
 2332 "ifndef", "elif" or "else" statement. The syntax is

2333  
 2334 "endif\n"  
 2335

#### 2336 4.3.14.8 Toggling example

2337  
 2338 Here is an example to show the workings of the toggling statements:  
 2339

2340 The "gensort" FDCC-set may be defined as:

```
2341
2342 LC_COLLATE
2343 ifdef BACKWARD
2344 order_start <LATIN>;forward;backward;forward;forward,position
2345 else
2346 order_start <LATIN>;forward;forward;forward;forward,position
2347 endif
2348 ....
2349 END LC_COLLATE
2350
```

2351 Then the following LC\_COLLATE category specification can use the "gensort" specification to create a new  
 2352 LC\_COLLATE category:

```
2353
2354 LC_COLLATE
2355 define BACKWARD
2356 copy "gensort"
2357 END LC_COLLATE
2358
```

2359 The example is explained as follows: The LC\_COLLATE category in the "gensort" FDCC-set uses the toggle  
 2360 "BACKWARD", and as "BACKWARD" is not set the second "order\_start" statement (all "forward") is used.

2361 In the second LC\_COLLATE category, the "BACKWARD" toggle is set before copying the first LC\_COLLATE  
 2362 category, and thus the first "order\_start" statement with 2nd level "backward" is used.

#### 2365 4.3.15 "i18n" LC\_COLLATE category

2366  
 2367 The "i18n" LC\_COLLATE category is defined as the following, which includes the  
 2368 tailorable template in ISO/IEC 14651.

```
2369 LC_COLLATE
2370
2371
2372 % Case collating symbols
2373 collating-symbol <RES-1>
2374 collating-symbol <BLK>
2375 collating-symbol <MIN> % SMALL
2376 collating-symbol <WIDE> % WIDE
2377 collating-symbol <COMPAT>
2378 collating-symbol <FONT>
```

2379	collating-symbol	<CIRCLE>	
2380	collating-symbol	<RES-2>	
2381	collating-symbol	<CAP>	% CAPITAL
2382	collating-symbol	<WIDECAP>	
2383	collating-symbol	<COMPATCAP>	
2384	collating-symbol	<FONTCAP>	
2385	collating-symbol	<CIRCLECAP>	
2386	collating-symbol	<HIRA-SMALL>	
2387	collating-symbol	<HIRA>	
2388	collating-symbol	<SMALL>	
2389	collating-symbol	<SMALL-NARROW>	
2390	collating-symbol	<KATA>	
2391	collating-symbol	<NARROW>	
2392	collating-symbol	<CIRCLE-KATA>	
2393	collating-symbol	<MNN>	
2394	collating-symbol	<MNS>	
2395	collating-symbol	<VERTICAL>	
2396	% Arabic forms		
2397	collating-symbol	<AINI>	
2398	collating-symbol	<AMED>	
2399	collating-symbol	<AFIN>	
2400	collating-symbol	<AISO>	
2401	%		
2402	collating-symbol	<NOBREAK>	
2403	collating-symbol	<SQUARED>	
2404	collating-symbol	<SQUARED CAP>	
2405	collating-symbol	<FRACTION>	
2406	collating-symbol	<BLANK>	
2407	collating-symbol	<CAPITAL-SMALL>	
2408	collating-symbol	<SMALL-CAPITAL>	
2409	collating-symbol	<BOTH>	
2410	% accents		
2411	collating-symbol	<LOWLINE>	% LOW LINE
2412	collating-symbol	<MACRO>	% MACRON
2413	collating-symbol	<OBLIK>	% STROKE
2414	collating-symbol	<AIGUT>	% ACUTE ACCENT
2415	collating-symbol	<GRAVE>	% GRAVE ACCENT
2416	collating-symbol	<BREVE>	% BREVE
2417	collating-symbol	<CIRCF>	% CIRCUMFLEX ACCENT
2418	collating-symbol	<CARON>	% CARON
2419	collating-symbol	<CRCL>	% RING ABOVE
2420	collating-symbol	<TREMA>	% DIAERESIS
2421	collating-symbol	<2AIGU>	% DOUBLE ACUTE ACCENT
2422	collating-symbol	<TILDE>	% TILDE
2423	collating-symbol	<POINT>	% DOT ABOVE
2424	collating-symbol	<CEDIL>	% CEDILLA
2425	collating-symbol	<OGONK>	% OGONEK
2426	collating-symbol	<OVERLINE>	% OVERLINE
2427	collating-symbol	<CROOK>	% HOOK ABOVE
2428	collating-symbol	<TONOS>	% VERTICAL LINE ABOVE
2429	collating-symbol	<D030E>	% DOUBLE VERTICAL LINE ABOVE
2430	collating-symbol	<2GRAV>	% DOUBLE GRAVE ACCENT
2431	collating-symbol	<D0310>	% CANDRABINDU
2432	collating-symbol	<BREVR>	% INVERTED BREVE
2433	collating-symbol	<D0312>	% TURNED COMMA ABOVE
2434	collating-symbol	<PSILI>	% COMMA ABOVE
2435	collating-symbol	<DASIA>	% REVERSED COMMA ABOVE
2436	collating-symbol	<D0315>	% COMMA ABOVE RIGHT
2437	collating-symbol	<D0316>	% GRAVE ACCENT BELOW
2438	collating-symbol	<D0317>	% ACUTE ACCENT BELOW
2439	collating-symbol	<D0318>	% LEFT TACK BELOW
2440	collating-symbol	<D0319>	% RIGHT TACK BELOW
2441	collating-symbol	<D031A>	% LEFT ANGLE ABOVE
2442	collating-symbol	<HORNU>	% HORN
2443	collating-symbol	<D031C>	% LEFT HALF RING BELOW
2444	collating-symbol	<D031D>	% UP TACK BELOW
2445	collating-symbol	<D031E>	% DOWN TACK BELOW
2446	collating-symbol	<D031F>	% PLUS SIGN BELOW
2447	collating-symbol	<D0320>	% MINUS SIGN BELOW
2448	collating-symbol	<PALCR>	% PALATALIZED HOOK BELOW
2449	collating-symbol	<RETCR>	% RETROFLEX HOOK BELOW
2450	collating-symbol	<POINS>	% DOT BELOW
2451	collating-symbol	<TREMS>	% DIAERESIS BELOW
2452	collating-symbol	<CRCLS>	% RING BELOW
2453	collating-symbol	<COMMS>	% COMMA BELOW
2454	collating-symbol	<D0329>	% VERTICAL LINE BELOW
2455	collating-symbol	<D032A>	% BRIDGE BELOW

2456	collating-symbol	<D032B>	% INVERTED DOUBLE ARCH BELOW
2457	collating-symbol	<D032C>	% CARON BELOW
2458	collating-symbol	<CIRCS>	% CIRCUMFLEX ACCENT BELOW
2459	collating-symbol	<BREVS>	% BREVE BELOW
2460	collating-symbol	<D032F>	% INVERTED BREVE BELOW
2461	collating-symbol	<TILDS>	% TILDE BELOW
2462	collating-symbol	<MACRS>	% MACRON BELOW
2463	collating-symbol	<D0333>	% DOUBLE LOW LINE
2464	collating-symbol	<TILDX>	% TILDE OVERLAY
2465	collating-symbol	<BARRE>	% SHORT STROKE OVERLAY
2466	collating-symbol	<D0336>	% LONG STROKE OVERLAY
2467	collating-symbol	<D0337>	% SHORT SOLIDUS OVERLAY
2468	collating-symbol	<CRCL2>	% RIGHT HALF RING BELOW
2469	collating-symbol	<D033A>	% INVERTED BRIDGE BELOW
2470	collating-symbol	<D033B>	% SQUARE BELOW
2471	collating-symbol	<D033C>	% SEAGULL BELOW
2472	collating-symbol	<D033D>	% X ABOVE
2473	collating-symbol	<D033E>	% VERTICAL TILDE
2474	collating-symbol	<D033F>	% DOUBLE OVERLINE
2475	collating-symbol	<PERIS>	% GREEK PERISPOMENI
2476	collating-symbol	<YPOGE>	% GREEK YPOGEGRAMMENI
2477	collating-symbol	<D0360>	% DOUBLE TILDE
2478	collating-symbol	<D0361>	% DOUBLE INVERTED BREVE
2479	collating-symbol	<DFE20>	% LIGATURE LEFT HALF
2480	collating-symbol	<DFE21>	% LIGATURE RIGHT HALF
2481	collating-symbol	<DFE22>	% DOUBLE TILDE LEFT HALF
2482	collating-symbol	<DFE23>	% DOUBLE TILDE RIGHT HALF
2483	collating-symbol	<D0483>	% CYRILLIC TITLO
2484	collating-symbol	<D0484>	% CYRILLIC PALATALIZATION
2485	collating-symbol	<D0485>	% CYRILLIC DASIA PNEUMATA
2486	collating-symbol	<D0486>	% CYRILLIC PSILI PNEUMATA
2487	collating-symbol	<SHEVA>	% HEBREW POINT SHEVA
2488	collating-symbol	<HTFSG>	% HEBREW POINT HATAF SEGOL
2489	collating-symbol	<HTFPPT>	% HEBREW POINT HATAF PATAH
2490	collating-symbol	<HTFQM>	% HEBREW POINT HATAF QAMATS
2491	collating-symbol	<HIRIQ>	% HEBREW POINT HIRIQ
2492	collating-symbol	<TSERE>	% HEBREW POINT TSERE
2493	collating-symbol	<SEGOL>	% HEBREW POINT SEGOL
2494	collating-symbol	<PATAH>	% HEBREW POINT PATAH
2495	collating-symbol	<QAMAT>	% HEBREW POINT QAMATS
2496	collating-symbol	<HOLAM>	% HEBREW POINT HOLAM
2497	collating-symbol	<QUBUT>	% HEBREW POINT QUBUTS
2498	collating-symbol	<DAGES>	% HEBREW POINT DAGESH OR MAPIQ
2499	collating-symbol	<RAPHE>	% HEBREW POINT RAFE
2500	collating-symbol	<SHINP>	% HEBREW POINT SHIN DOT
2501	collating-symbol	<SINPT>	% HEBREW POINT SIN DOT
2502	collating-symbol	<VARIKA>	% HEBREW POINT JUDEO-SPANISH VARIKA
2503	collating-symbol	<FATHATAN>	% ARABIC FATHATAN
2504	collating-symbol	<DAMMATAN>	% ARABIC DAMMATAN
2505	collating-symbol	<KASRATAN>	% ARABIC KASRATAN
2506	collating-symbol	<FATHA>	% ARABIC FATHA
2507	collating-symbol	<DAMMA>	% ARABIC DAMMA
2508	collating-symbol	<KASRA>	% ARABIC KASRA
2509	collating-symbol	<SHADDA>	% ARABIC SHADDA
2510	collating-symbol	<SUKUN>	% ARABIC SUKUN
2511	collating-symbol	<SUPERALEF>	% ARABIC LETTER SUPERScript ALEF
2512	collating-symbol	<D06D6>	% ARABIC SMALL HIGH LIGATURE SAD WITH LAM WITH ALEF MAKSURA
2513	collating-symbol	<D06D7>	% ARABIC SMALL HIGH LIGATURE QAF WITH LAM WITH ALEF MAKSURA
2514	collating-symbol	<D06D8>	% ARABIC SMALL HIGH MEEM INITIAL FORM
2515	collating-symbol	<D06D9>	% ARABIC SMALL HIGH LAM ALEF
2516	collating-symbol	<D06DA>	% ARABIC SMALL HIGH JEEM
2517	collating-symbol	<D06DB>	% ARABIC SMALL HIGH THREE DOTS
2518	collating-symbol	<D06DC>	% ARABIC SMALL HIGH SEEN
2519	collating-symbol	<D06E1>	% ARABIC SMALL HIGH DOTLESS HEAD OF KHAH
2520	collating-symbol	<D06E2>	% ARABIC SMALL HIGH MEEM ISOLATED FORM
2521	collating-symbol	<D06E3>	% ARABIC SMALL LOW SEEN
2522	collating-symbol	<AMADD>	% ARABIC SMALL HIGH MADDA
2523	collating-symbol	<D06E7>	% ARABIC SMALL HIGH YEH
2524	collating-symbol	<D06E8>	% ARABIC SMALL HIGH NOON
2525	collating-symbol	<D06ED>	% ARABIC SMALL LOW MEEM
2526	collating-symbol	<D093C>	% DEVANAGARI SIGN NUKTA
2527	collating-symbol	<D0951>	% DEVANAGARI STRESS SIGN UDATTA
2528	collating-symbol	<D0952>	% DEVANAGARI STRESS SIGN ANUDATTA
2529	collating-symbol	<D0953>	% DEVANAGARI GRAVE ACCENT
2530	collating-symbol	<D0954>	% DEVANAGARI ACUTE ACCENT
2531	collating-symbol	<D09BC>	% BENGALI SIGN NUKTA
2532	collating-symbol	<D0A3C>	% GURMUKHI SIGN NUKTA
2533	collating-symbol	<D0ABC>	% GUJARATI SIGN NUKTA
2534	collating-symbol	<D0B3C>	% ORIYA SIGN NUKTA

2535	collating-symbol	<D0E48>	% THAI CHARACTER MAI EK
2536	collating-symbol	<D0E49>	% THAI CHARACTER MAI THO
2537	collating-symbol	<D0E4A>	% THAI CHARACTER MAI TRI
2538	collating-symbol	<D0E4B>	% THAI CHARACTER MAI CHATAWA
2539	collating-symbol	<D0EC8>	% LAO TONE MAI EK
2540	collating-symbol	<D0EC9>	% LAO TONE MAI THO
2541	collating-symbol	<D0ECA>	% LAO TONE MAI TI
2542	collating-symbol	<D0ECB>	% LAO TONE MAI CATAWA
2543	collating-symbol	<D0F39>	% TIBETAN MARK TSA -PHRU
2544	collating-symbol	<D0F3E>	% TIBETAN SIGN YAR TSHES
2545	collating-symbol	<D0F3F>	% TIBETAN SIGN MAR TSHES
2546	collating-symbol	<D302A>	% IDEOGRAPHIC LEVEL TONE MARK
2547	collating-symbol	<D302B>	% IDEOGRAPHIC RISING TONE MARK
2548	collating-symbol	<D302C>	% IDEOGRAPHIC DEPARTING TONE MARK
2549	collating-symbol	<D302D>	% IDEOGRAPHIC ENTERING TONE MARK
2550	collating-symbol	<D302E>	% HANGUL SINGLE DOT TONE MARK
2551	collating-symbol	<D302F>	% HANGUL DOUBLE DOT TONE MARK
2552	collating-symbol	<KNVCE>	% KATAKANA-HIRAGANA VOICED SOUND MARK
2553	collating-symbol	<KNSMV>	% KATAKANA-HIRAGANA SEMI-VOICED SOUND MARK
2554	collating-symbol	<D20D0>	% LEFT HARPOON ABOVE
2555	collating-symbol	<D20D1>	% RIGHT HARPOON ABOVE
2556	collating-symbol	<D20D2>	% LONG VERTICAL LINE OVERLAY
2557	collating-symbol	<D20D3>	% SHORT VERTICAL LINE OVERLAY
2558	collating-symbol	<D20D4>	% ANTICLOCKWISE ARROW ABOVE
2559	collating-symbol	<D20D5>	% CLOCKWISE ARROW ABOVE
2560	collating-symbol	<D20D6>	% LEFT ARROW ABOVE
2561	collating-symbol	<D20D7>	% RIGHT ARROW ABOVE
2562	collating-symbol	<D20D8>	% RING OVERLAY
2563	collating-symbol	<D20D9>	% CLOCKWISE RING OVERLAY
2564	collating-symbol	<D20DA>	% ANTICLOCKWISE RING OVERLAY
2565	collating-symbol	<D20DB>	% THREE DOTS ABOVE
2566	collating-symbol	<D20DC>	% FOUR DOTS ABOVE
2567	collating-symbol	<D20DD>	% ENCLOSING CIRCLE
2568	collating-symbol	<D20DE>	% ENCLOSING SQUARE
2569	collating-symbol	<D20DF>	% ENCLOSING DIAMOND
2570	collating-symbol	<D20E0>	% ENCLOSING CIRCLE BACKSLASH
2571	collating-symbol	<D20E1>	% LEFT RIGHT ARROW ABOVE
2572	collating-symbol	<NEGATIVE>	
2573	collating-symbol	<SANSSERIF>	
2574	collating-symbol	<NEGSANSSERIF>	
2575	collating-symbol	<ARABIC>	
2576	collating-symbol	<EXTARABIC>	
2577	collating-symbol	<NAGAR>	
2578	collating-symbol	<BENGL>	
2579	collating-symbol	<BENGALINUMERATOR>	
2580	collating-symbol	<GURMU>	
2581	collating-symbol	<GUJAR>	
2582	collating-symbol	<ORIYA>	
2583	collating-symbol	<TAMIL>	
2584	collating-symbol	<TELGU>	
2585	collating-symbol	<KNNDA>	
2586	collating-symbol	<MALAY>	
2587	collating-symbol	<SINHALA>	
2588	collating-symbol	<THAII>	
2589	collating-symbol	<LAAOO>	
2590	collating-symbol	<BODKA>	
2591	collating-symbol	<CJKVS>	
2592	collating-symbol	<S0200>..<<S1100>	% 0x0200..0x1100
2593			
2594	collating-symbol	<S4E00>..<<S9FA5>	% Symbols for Han
2595			
2596	collating-symbol	<SAC00>..<<SD7A3>	% Symbols for Hangul
2597			
2598	collating-symbol	<SFA0E>..<<SFA29>	% Symbols for Compatibility Han
2599			
2600	% equivalences		
2601	symbol-equivalence	<NONE>	<BLANK>
2602	symbol-equivalence	<CAPITAL>	<CAP>
2603	symbol-equivalence	<MACRON>	<MACRO>
2604	symbol-equivalence	<STROKE>	<OBLIK>
2605	symbol-equivalence	<ACUTE>	<AIGUT>
2606	symbol-equivalence	<CIRCUMFLEX>	<CIRCF>
2607	symbol-equivalence	<RING>	<CRCLE>
2608	symbol-equivalence	<DIAERESIS>	<TREMA>
2609	symbol-equivalence	<DOT>	<POINT>
2610	symbol-equivalence	<CEDILLA>	<CEDIL>
2611	symbol-equivalence	<OGONEK>	<OGONK>

```

2612 symbol-equivalence <HOOK> <CROOK>
2613 symbol-equivalence <HORN> <HORNU>
2614 symbol-equivalence <DOT-BELOW> <POINS>
2615
2616 order_start <Latin>;forward;backward;forward;forward,position
2617
2618 % Copy the template from ISO/IEC 14651
2619 copy "isol4651t1"
2620
2621 order_end
2622
2623 END LC_COLLATE
2624

```

#### 4.4 LC\_MONETARY

The LC\_MONETARY category defines the rules and symbols that shall be used to format monetary numeric information. The operands are strings. For some keywords, the strings can contain only integers. More than one set of monetary values may be provided, and for each set a period of validity and conversion rate may be given. Keywords that are not provided, string values set to the empty string "", or integer keywords set to -1, shall be used to indicate that the value is unspecified, and then no default is taken. The following keywords shall be defined:

2635	<b>copy</b>	Specify the name of an existing FDCC-set to be used as the source for the definition of this category. If this keyword is specified, no other keyword shall be specified.
2637	<b>valid_from</b>	One or more integers separated by semicolons, representing a Gregorian date in the form YYYYMMDD, specifying the beginning date (inclusive) of the validity of a currency. The position of the integer in the list corresponds to the position of operands in other keywords in the LC_MONETARY category. The currencies should be ordered in terms of validity dates, and for each validity period with the currency that the amounts are stored in first. If not specified, it is taken to be the beginning of time.
2647	<b>valid_to</b>	One or more integers separated by semicolons, representing a Gregorian date in the form YYYYMMDD, specifying the end date (inclusive) of the validity of a currency. If not specified, it is taken to be the end of time.
2651	<b>conversion_rate</b>	one or more pairs of integers separated by a <semicolon> specifying the fixed conversion rate between the currency in question and the first valid currency for the period. If the currency is not the first valid currency for the period in question, the first integer is for multiplying the first currency, and the second for dividing this result to get the amount in the currency in question. Each pair of integers are separated by a <slash>. The default value is "1/100". This keyword is optional.
2660	<b>int_curr_symbol</b>	One or more strings separated by semicolons that shall be used as the international currency symbols. Each operand shall be a four character string, with the first three characters containing the alphabetic international currency symbol in accordance with those specified in ISO 4217 (Codes for the representation of currencies and funds). The fourth character shall be the character used to separate the international

2667		currency symbol from the monetary quantity. The keyword shall be specified, unless the "copy" keyword is used.
2668		
2669	<b>currency_symbol</b>	One or more strings separated by semicolons that shall be used as the local currency symbol.
2670		
2671	<b>mon_decimal_point</b>	The operand is a string containing the symbol that shall be used as the decimal delimiter in monetary formatted quantities. In contexts where other standards limit the "mon_decimal_point" to a single byte, the result of specifying a multibyte operand is unspecified. The keyword shall be specified, unless the "copy" keyword is used.
2672		
2673		
2674		
2675		
2676		
2677	<b>mon_thousands_sep</b>	The operand is a string containing the symbol that shall be used as a separator for groups of digits to the left of the decimal delimiter in formatted monetary quantities. In contexts where other standards limit the "mon_thousands_sep" to a single byte, the result of specifying a multibyte operand is unspecified. The keyword shall be specified, unless the "copy" keyword is used.
2678		
2679		
2680		
2681		
2682		
2683		
2684	<b>mon_grouping</b>	Define the size of each group of digits in formatted monetary quantities. The operand is a sequence of integers separated by semicolons. Each integer specifies the number of digits in each group, with the initial integer defining the size of the group immediately preceding the decimal delimiter, and the following integers defining the preceding groups. If the last integer is not -1, then the size of the previous group (if any) shall be repeatedly used for the remainder of the digits. If the last integer is -1, then no further grouping shall be performed. The keyword shall be specified, unless the "copy" keyword is used.
2685		
2686		
2687		
2688		
2689		
2690		
2691		
2692		
2693		
2694		
2695	<b>positive_sign</b>	A string that shall be used to indicate a nonnegative-valued formatted monetary quantity. The keyword shall be specified, unless the "copy" keyword is used.
2696		
2697		
2698	<b>negative_sign</b>	A string that shall be used to indicate a negative-valued formatted monetary quantity. The keyword shall be specified, unless the "copy" keyword is used.
2699		
2700		
2701	<b>int_frac_digits</b>	One or more integers separated by semicolons, representing the number of fractional digits (those to the right of the decimal delimiter) to be written in a formatted monetary quantity using int_curr_symbol. The keyword shall be specified, unless the "copy" keyword is used.
2702		
2703		
2704		
2705		
2706	<b>frac_digits</b>	One or more integers separated by semicolons, representing the number of fractional digits (those to the right of the decimal delimiter) to be written in a formatted monetary quantity using "currency_symbol". The keyword shall be specified, unless the "copy" keyword is used.
2707		
2708		
2709		
2710		
2711	<b>p_cs_precedes</b>	One or more integers separated by semicolons, set to 1 if the "currency_symbol" precedes the value for a nonnegative formatted monetary quantity, and set to 0 if the symbol succeeds the value. The keyword shall be specified, unless the "copy" keyword is used.
2712		
2713		
2714		
2715		
2716	<b>p_sep_by_space</b>	One or more integers separated by semicolons, set to 0 if no



2717		space separates the "currency_symbol" from the value for a
2718		nonnegative formatted monetary quantity, set to 1 if a space
2719		separates the symbol from the value, and set to 2 if a space
2720		separates the symbol and the sign string, if adjacent. The
2721		keyword shall be specified, unless the "copy" keyword is
2722		used.
2723	<b>n_cs_precedes</b>	One or more integers separated by semicolons, set to 1 if the
2724		"currency_symbol" precedes the value for a negative
2725		formatted monetary quantity, and set to 0 if the symbol
2726		succeeds the value. The keyword shall be specified, unless
2727		the "copy" keyword is used.
2728	<b>n_sep_by_space</b>	One or more integers separated by semicolons, set to 0 if no
2729		space separates the "currency_symbol" from the value for a
2730		negative formatted monetary quantity, set to 1 if a space
2731		separates the symbol from the value, and set to 2 if a space
2732		separates the symbol and the sign string, if adjacent. The
2733		keyword shall be specified, unless the "copy" keyword is
2734		used.
2735	<b>int_p_cs_precedes</b>	One or more integers separated by semicolons; set to 1 if the
2736		"int_curr_symbol" precedes the value for a nonnegative
2737		formatted monetary quantity, and set to 0 if the symbol
2738		succeeds the value. If not specified, the value of
2739		"p_cs_precedes" is taken.
2740	<b>int_p_sep_by_space</b>	One or more integers separated by semicolons; set to 0 if no
2741		space separates the "int_curr_symbol" from the value for a
2742		nonnegative formatted monetary quantity, set to 1 if a space
2743		separates the symbol from the value, and set to 2 if a space
2744		separates the symbol and the sign string, if adjacent. If not
2745		specified, the value of "p_sep_by_space" is taken.
2746	<b>int_n_cs_precedes</b>	One or more integers separated by semicolons; set to 1 if the
2747		"int_curr_symbol" precedes the value for a negative
2748		formatted monetary quantity, and set to 0 if the symbol
2749		succeeds the value. If not specified, the value of
2750		"n_cs_precedes" is taken.
2751	<b>int_n_sep_by_space</b>	One or more integers separated by semicolons; set to 0 if no
2752		space separates the "int_curr_symbol" from the value for a
2753		negative formatted monetary quantity, set to 1 if a space
2754		separates the symbol from the value, and set to 2 if a space
2755		separates the symbol and the sign string, if adjacent. If not
2756		specified, the value of "n_sep_by_space" is taken.
2757	<b>p_sign_posn</b>	One or more integers separated by semicolons, set to a value
2758		indicating the positioning of the "positive_sign" for a
2759		nonnegative formatted monetary quantity using the
2760		"currency_symbol". The following integer values shall be
2761		defined:
2762		
2763		0    Parentheses enclose the quantity and the
2764		"currency_symbol".
2765		1    The sign string precedes the quantity and the
2766		"currency_symbol".

2767		2	The sign string succeeds the quantity and the "currency_symbol".
2768			
2769		3	The sign string immediately precedes the "currency_symbol".
2770			
2771		4	The sign string immediately succeeds the "currency_symbol".
2772			
2773			The keyword shall be specified, unless the "copy" keyword is used.
2774			
2775			
2776	<b>n_sign_posn</b>		One or more integers separated by semicolons, set to a value indicating the positioning of the "negative_sign" for a negative formatted monetary quantity using the "currency_symbol". The following integer values shall be defined:
2777			
2778			
2779			
2780			
2781			
2782		0	Parentheses enclose the quantity and the "currency_symbol".
2783			
2784		1	The sign string precedes the quantity and the "currency_symbol".
2785			
2786		2	The sign string succeeds the quantity and the "currency_symbol".
2787			
2788		3	The sign string immediately precedes the "currency_symbol".
2789			
2790		4	The sign string immediately succeeds the "currency_symbol".
2791			
2792			The keyword shall be specified, unless the "copy" keyword is used.
2793			
2794			
2795	<b>int_p_sign_posn</b>		One or more integers separated by semicolons, set to a value indicating the positioning of the "positive_sign" for a nonnegative formatted international monetary quantity. The following integer values shall be defined:
2796			
2797			
2798			
2799			
2800		0	Parentheses enclose the quantity and the "int_curr_symbol".
2801			
2802		1	The sign string precedes the quantity and the "int_curr_symbol".
2803			
2804		2	The sign string succeeds the quantity and the "int_curr_symbol".
2805			
2806		3	The sign string immediately precedes the "int_curr_symbol".
2807			
2808		4	The sign string immediately succeeds the "int_curr_symbol".
2809			
2810			If no "int_p_sign_posn" is present the value of the "p_sign_posn" is taken.
2811			
2812			
2813	<b>int_n_sign_posn</b>		One or more integers separated by semicolons, set to a value indicating the positioning of the "negative_sign" for a negative formatted international monetary quantity. The following integer values shall be defined:
2814			
2815			
2816			

- 2817 0 Parentheses enclose the quantity and the  
 2818 "int\_curr\_symbol".  
 2819 1 The sign string precedes the quantity and the  
 2820 "int\_curr\_symbol".  
 2821 2 The sign string succeeds the quantity and the  
 2822 "int\_curr\_symbol".  
 2823 3 The sign string immediately precedes the  
 2824 "int\_curr\_symbol".  
 2825 4 The sign string immediately succeeds the  
 2826 "int\_curr\_symbol".  
 2827 If no "int\_n\_sign\_posn" is present the value of the  
 2828 "n\_sign\_posn" is taken.  
 2829

2830 The "i18n" FDCC-set is defined as follows for the LC\_MONETARY category.

```

2831 LC_MONETARY
2832 % This is the 14652 i18n fdcc-set definition for
2833 % the LC_MONETARY category.
2834 %
2835 %
2836 int_curr_symbol      ""
2837 currency_symbol     ""
2838 mon_decimal_point    ""
2839 mon_thousands_sep   ""
2840 mon_grouping         -1
2841 positive_sign        ""
2842 negative_sign        ""
2843 int_frac_digits      -1
2844 frac_digits          -1
2845 p_cs_precedes        -1
2846 p_sep_by_space       -1
2847 n_cs_precedes        -1
2848 n_sep_by_space       -1
2849 p_sign_posn          -1
2850 n_sign_posn          -1
2851 %
2852 END LC_MONETARY
  
```

#### 2855 4.5 LC\_NUMERIC

2856  
 2857 The LC\_NUMERIC category defines the rules and symbols that shall be used to format  
 2858 nonmonetary numeric information. The operands are strings. For some keywords, the  
 2859 strings only can contain integers. Keywords that are not provided, string values set to the  
 2860 empty string (""), or integer keywords set to -1, shall be used to indicate that the value is  
 2861 unspecified. The following keywords shall be defined:

- 2862  
 2863 **copy** Specify the name of an existing FDCC-set to be used as the  
 2864 source for the definition of this category. If this keyword is  
 2865 specified, no other keyword shall be specified.  
 2866 **decimal\_point** The operand is a string containing the symbol that shall be used  
 2867 as the decimal delimiter in numeric, nonmonetary formatted  
 2868 quantities. This keyword cannot be omitted and cannot be set to  
 2869 the empty string. In contexts where other standards limit the  
 2870 decimal point to a single byte, the result of specifying a mul-  
 2871 ti-byte operand is unspecified.  
 2872 **thousands\_sep** The operand is a string containing the symbol that shall be used  
 2873 as a separator for groups of digits to the left of the decimal  
 2874 delimiter in numeric, nonmonetary formatted monetary quan-  
 2875 tities. In contexts where other standards limit the

2876 "thousands\_sep" to a single byte, the result of specifying a  
 2877 multibyte operand is unspecified.  
 2878 **grouping** Define the size of each group of digits in formatted non-  
 2879 monetary quantities. The operand is a sequence of integers  
 2880 separated by semicolons. Each integer specifies the number of  
 2881 digits in each group, with the initial integer defining the size of  
 2882 the group immediately preceding the decimal delimiter, and the  
 2883 following integers defining the preceding groups. If the last  
 2884 integer is not -1, then the size of the previous group (if any)  
 2885 shall be repeatedly used for the remainder of the digits. If the  
 2886 last integer is -1, then no further grouping shall be performed.  
 2887

2888 The "i18n" FDCC-set is for the LC\_NUMERIC category:

```
2889 LC_NUMERIC
2890 % This is the 14652 i18n fdcc-set definition for
2891 % the LC_NUMERIC category.
2892 %
2893 %
2894 decimal_point    ""
2895 thousands_sep    ""
2896 grouping         -1
2897 %
2898 END LC_NUMERIC
2899
```

2900

#### 2901 4.6 LC\_TIME

2902

2903 The LC\_TIME category defines the rules and symbols that shall be used to format date  
 2904 and time information. The following keywords shall be defined:

2905

2906 **copy** Specify the name of an existing FDCC-set to be used as the source  
 2907 for the definition of this category. If this keyword is specified, no  
 2908 other keyword shall be specified.

2909 **abday** Define the abbreviated weekday names for calendar systems with  
 2910 weeks of constant length, to be referenced by the %a field descriptor.  
 2911 The length of the week and a gregorian date for the first weekday is  
 2912 defined by the "week" keyword. The operand shall consist of  
 2913 semicolon-separated strings. The first string shall be the abbreviated  
 2914 name of the day corresponding to the first day of the week (default  
 2915 Sunday), the second the abbreviated name of the day corresponding  
 2916 to the second day of the week (default Monday), and so on.

2917 **day** Define the full weekday names for calendar systems with weeks of  
 2918 constant length, to be referenced by the %A field descriptor. The  
 2919 length of the week and a gregorian date for the first weekday is  
 2920 defined by the "week" keyword. The operand shall consist of  
 2921 semicolon-separated strings. The first string shall be the full name of  
 2922 the day corresponding to the first day of the week (default Sunday),  
 2923 the second the full name of the day corresponding to the second day  
 2924 of the week (default Monday), and so on.

2925 **week** Shall be used to define the number of days in a week, which is the  
 2926 first weekday - the first weekday has the value 1, and which week is  
 2927 to be considered the first in a year. The first operand is an integer  
 2928 specifying the number of days in the week, The second operand is an  
 2929 integer specifying the gregorian date in the format YYYYMMDD

2930		with a leading <hyphen-minus> if before Christ. The third operand is
2931		an integer specifying the weekday number to be contained in the first
2932		week of the year. If the keyword is not specified the values are taken
2933		as 7, 19971130 (a Sunday), and 7 (Saturday), respectively. ISO 8601
2934		conforming applications should use the values 7, 19971201 (a
2935		Monday), and 4 (Thursday), respectively.
2936	<b>abmon</b>	Define the abbreviated month names, to be referenced by the %b
2937		field descriptor. The operand shall consist of twelve or thirteen
2938		semicolon-separated strings. The first string shall be the abbreviated
2939		name of the first month of the year (January), the second the
2940		abbreviated name of the second month, and so on.
2941	<b>mon</b>	Define the full month names, to be referenced by the %B field
2942		descriptor. The operand shall consist of twelve or thirteen semicolon-
2943		separated strings. The first string shall be the full name of the first
2944		month of the year (January), the second the full name of the second
2945		month, and so on.
2946	<b>d_t_fmt</b>	Define the appropriate date and time representation, to be referenced
2947		by the %c field descriptor. The operand shall consist of a string, and
2948		can contain any combination of characters and field descriptors. In
2949		addition, the string can contain escape sequences defined in Table 3.
2950	<b>d_fmt</b>	Define the appropriate date representation, to be referenced by the
2951		%x field descriptor. The operand shall consist of a string, and can
2952		contain any combination of characters and field descriptors. In
2953		addition, the string can contain escape sequences defined in Table 3.
2954	<b>t_fmt</b>	Define the appropriate time representation, to be referenced by the
2955		%X field descriptor. The operand shall consist of a string, and can
2956		contain any combination of characters and field descriptors. In
2957		addition, the string can contain escape sequences defined in Table 3.
2958	<b>am_pm</b>	Define the appropriate representation of the ante meridiem and post
2959		meridiem strings, to be referenced by the %p field descriptor. The
2960		operand shall consist of two strings, separated by a semicolon. The
2961		first string shall represent the antemeridiem designation, the last
2962		string the postmeridiem designation. The keyword is optional. If
2963		unspecified, the %p field descriptor shall refer to the empty string.
2964	<b>t_fmt_ampm</b>	Define the appropriate time representation in the 12-hour clock
2965		format with "am_pm", to be referenced by the %r field descriptor.
2966		The operand shall consist of a string and can contain any
2967		combination of characters and field descriptors. If the string is empty,
2968		the 12-hour format is not supported in the FDCC-set.
2969	<b>era</b>	Shall be used to define alternate Eras, corresponding to the %E field
2970		descriptor modifier. The format of the operand is unspecified, but
2971		shall support the definition of the %EC and %Ey field descriptors,
2972		and may also define the "era_year" format (%EY).
2973	<b>era_year</b>	Shall be used to define the format of the year in alternate Era format,
2974		corresponding to the %EY field descriptor.
2975	<b>era_d_fmt</b>	Shall be used to define the format of the date in alternate Era
2976		notation, corresponding to the %Ex field descriptor.
2977	<b>alt_digits</b>	Shall be used to define alternate symbols for digits, corresponding to
2978		the %O field descriptor modifier. The operand shall consist of
2979		semicolon-separated strings. The first string shall be the alternate

2980		symbol corresponding with zero, the second string the symbol
2981		corresponding with one, and so on. Up to 100 alternate symbol
2982		strings can be specified. The %O modifier indicates that the string
2983		corresponding to the value specified via the field descriptor shall be
2984		used instead of the value.
2985	<b>first_weekday</b>	Shall be used to define the first day to be displayed, for example in a
2986		calendar display utility. The operand is an integer specifying the day
2987		number (1 = first) according to the information specified with the
2988		"day" keyword. The keyword may be omitted, and then the value 1 is
2989		taken, corresponding to Sunday for a week beginning Sunday, or to
2990		Monday for a week beginning Monday.
2991	<b>first_workday</b>	Shall be used to define the first workday as an integer according to
2992		the day numbering specified with the "week" keyword.
2993	<b>cal_direction</b>	Shall be used to define the direction of the display of dates, for
2994		example in a calendar display utility. The operand is an integer, and
2995		the following values are defined:
2996		1 left-right from top
2997		2 top-down from left
2998		3 right-left from top
2999		The keyword may be omitted, and then the value 1 is taken.
3000	<b>timezone</b>	Shall be used to define a set of timezones, each defined by a string.
3001		In the following the characters <, >, [ and ] are used as
3002		metacharacters. Only characters with a visible glyph from the
3003		portable character set may be used, except in the <std> and <dst>
3004		fields. The syntax of the string is:
3005		
3006		<std><offset><dst>[<offset>][,<rule>[,<rule>...]]
3007		
3008		where
3009		
3010		<std> and <dst> Indicates no less than three, nor more than 10
3011		characters that are the designation for the
3012		standard <std> or summer <dst> time zone.
3013		only <std> is required; if <dst> is missing, then
3014		summer time does not apply in this category.
3015		Upper- and lowercase letters are explicitly
3016		allowed. Any characters except a leading colon
3017		<:> or digits, the comma <,>, the minus <->,
3018		the plus <+>, and the null character are
3019		permitted to appear in these fields, but their
3020		meaning is unspecified.
3021		<offset> Indicates the value one must add to the local
3022		time to arrive at the Coordinated Universal
3023		Time. The <offset> has the form:
3024		
3025		hh[:mm[:ss]]
3026		
3027		The minutes (mm) and seconds (ss) are
3028		optional. The hour (hh) shall be required and
3029		may be a single digit. The <offset> following

3030		<std> shall be required. If no <offset> follows
3031		<dst>, summer time is assumed to be one hour
3032		ahead of standard time. One or more digits may
3033		be used; the value is always interpreted as a
3034		decimal number. The hour shall be between
3035		zero and 24, and the minutes (and seconds) - if
3036		present - shall be between zero and 59. If
3037		preceded by a "-", the time zone shall be east
3038		of the Prime Meridian; otherwise it shall be
3039		west of (which may be indicated by an optional
3040		preceding "+").
3041	<rule>	Indicates when to change to and back from
3042		summer time. The <rule> has the form:
3043		<date>[/<time>/<year>],<date>[/<time>
3044		/<year>]
3045		where the first <date> describes when the
3046		change from standard time to summer time
3047		occurs, and the second <date> describes when
3048		the change back happens. Each <time> field
3049		describes when, in current local time, the
3050		change to the other time is made. The first
3051		<year> field defines the beginning of the
3052		validity of this rule, and the second <year>
3053		field defines the end of the validity of the rule.
3054		A number of rules may be given.
3055		
3056		The format of <date> shall be one of the
3057		following:
3058		
3059	J<n>	The Julian day <n> (1 <= n
3060		<= 365) Leap years shall not
3061		be counted. That is, in all
3062		years - including leap years -
3063		February 28 is day 59 and
3064		March 1 is day 60. It is
3065		impossible to explicitly refer
3066		to the occasional February 29.
3067	<n>	The zero-based Julian day (0
3068		<= n <= 365). Leap years
3069		shall be counted and it is
3070		possible to refer to February
3071		29.
3072	M<m>.<n>.<d>	
3073		the <d>th day (0 <= d <= 7)
3074		of week <n> of month <m> (1
3075		<= n <= 5, 1 <= m <= 12,
3076		where week 5 means "the last
3077		<d> day in month <m>"
3078		which may occur in either the
3079		fourth or fifth week). Week 1

3080 is the first week in which the  
 3081 <d>th day occurs. Day zero  
 3082 and day seven is Sunday.

3083  
 3084 The <time> has the same format as <offset>  
 3085 except that no leading sign ("- or "+) shall be  
 3086 allowed. The default, if <time> is not given,  
 3087 shall be "02:00:00".

3088  
 3089 The <year> has the format YYYY.  
 3090

#### 3091 4.6.1 Date Field Descriptors

3092  
 3093 The LC\_TIME category defines the interpretation of a number of field descriptors. The  
 3094 field descriptors are also available in the definitions with the following LC\_TIME  
 3095 keywords: "d\_t\_fmt", "d\_fmt", "t\_fmt", "t\_fmt\_ampm", "era", and "era\_d\_fmt". A field  
 3096 descriptor may not be used with the LC\_TIME keywords defining it.  
 3097

3098 Table 3: Escape sequences for the date field

3099		
3100	%a	FDCC-set's abbreviated weekday name.
3101	%A	FDCC-set's full weekday name.
3102	%b	FDCC-set's abbreviated month name.
3103	%B	FDCC-set's full month name.
3104	%c	FDCC-set's appropriate date and time representation.
3105	%C	Century (a year divided by 100 and truncated to integer) as decimal 3106 number (00-99).
3107	%d	Day of the month as a decimal number (01-31).
3108	%D	Date in the format mm/dd/yy.
3109	%e	Day of the month as a decimal number (1-31 in at two-digit field with 3110 leading <space> fill).
3111	%F	is replaced by the date in the format YYYY-MM-DD (ISO 8601 format)
3112	%h	A synonym for %b.
3113	%H	Hour (24-hour clock) as a decimal number (00-23).
3114	%I	Hour (12-hour clock) as a decimal number (01-12).
3115	%j	Day of the year as a decimal number (001-366).
3116	%m	Month as a decimal number (01-13).
3117	%M	Minute as a decimal number (00-59).
3118	%n	A <newline> character.
3119	%p	FDCC-set's equivalent of either AM or PM.
3120	%r	12-hour clock time (01-12) using the AM/PM notation.
3121	%S	Seconds as a decimal number (00-61).
3122	%t	A <tab> character.
3123	%T	24-hour clock time in the format HH:MM:SS.
3124	%u	Weekday as a decimal number (1(Monday)-7).
3125	%U	Week number of the year (Sunday as the first day of the week) as a 3126 decimal number (00-53).
3127	%v	Week number of the year as a decimal number with two digits including a 3128 possible leading zero, according to "week" keyword.
3129	%V	Week of the year (Monday as the first day of the week) as a decimal



3130		number (01-53). The method for determining the week number shall be as
3131		specified by ISO 8601.
3132	%w	Weekday as a decimal number (0(Sunday)-6).
3133	%W	Week number of the year (Monday as the first day of the week) as a
3134		decimal number (00-53).
3135	%x	FDCC-set's appropriate date representation.
3136	%X	FDCC-set's appropriate time representation.
3137	%y	Year (offset from %C) as a decimal number (00-99).
3138	%Y	Year with century as a decimal number.
3139	%Z	Time-zone name, or no characters if no time zone is determinable.
3140	%%	A <percent-sign> character.

3141

#### 3142 **4.6.2 Modified Field Descriptors**

3143

3144 Some field descriptors can be modified by the E and O modifier characters to indicate a  
 3145 different format or specification as specified in the LC\_TIME FDCC-set description. If the  
 3146 corresponding keyword (see "era", "era\_year", "era\_d\_fmt", and "alt\_digits") is not  
 3147 specified for the current FDCC-set, the unmodified field descriptor value shall be used.

3148

3149	%Ec	FDCC-set's alternate date and time representation.
3150	%EC	The name of the base year (period) in the FDCC-set's alternate represen-
3151		tation.
3152	%Ex	FDCC-set's alternate date representation.
3153	%Ey	Offset from %EC (year only) in the FDCC-set's alternate representation.
3154	%EY	Full alternate year representation.
3155	%Od	Day of month using the FDCC-set's alternate numeric symbols.
3156	%Oe	Day of month using the FDCC-set's alternate numeric symbols.
3157	%Of	Weekday as a decimal number according to alt_day (1 is first day).
3158	%OH	Hour (24-hour clock) using the FDCC-set's alternate numeric symbols.
3159	%OI	Hour (12-hour clock) using the FDCC-set's alternate numeric symbols.
3160	%Om	Month using the FDCC-set's alternate numeric symbols.
3161	%OM	Minutes using the FDCC-set's alternate numeric symbols.
3162	%OS	Seconds using the FDCC-set's alternate numeric symbols.
3163	%Ou	Weekday as a number in the alternate representation of the FDCC-set
3164		(Monday=1).
3165	%OU	Week number of the year (Sunday as the first day of the week) using the
3166		FDCC-set's alternate numeric symbols.
3167	%OV	Week number of the year (Monday as the first day of the week, ISO 8601
3168		rules) using the alternate numeric symbols of the FDCC-set.
3169	%Ow	Weekday as number in the FDCC-set's alternate representation
3170		(Sunday=0).
3171	%OW	Week number of the year (Monday as the first day of the week) using the
3172		FDCC-set's alternate numeric symbols.
3173	%Oy	Year (offset from %C) in alternate representation.

3174

#### 3175 **4.6.3 "i18n" LC\_TIME category**

3176

3177 The "i18n" LC\_TIME category is (following ISO 8601):

3178

```
3179     LC_TIME
3180     % This is the ISO/IEC 14652 "i18n" definition for
3181     % the LC_TIME category.
```

```

3182 %
3183 % Weekday and week numbering according to ISO 8601
3184 abday "<1>";"<2>";"<3>";"<4>";"<5>";"<6>;<7>"
3185 day "<1>";"<2>";"<3>";"<4>";"<5>";"<6>;<7>"
3186 week 7;19971201;4
3187 abmon "<0><1>";"<0><2>";"<0><3>";"<0><4>";"<0><5>";"<0><6>";/
3188 "<0><7>";"<0><8>";"<0><9>";"<1><0>";"<1><1>";"<1><2>"
3189 mon "<0><1>";"<0><2>";"<0><3>";"<0><4>";"<0><5>";"<0><6>";/
3190 "<0><7>";"<0><8>";"<0><9>";"<1><0>";"<1><1>";"<1><2>"
3191 am_pm "";"
3192 % Date formats following ISO 8601
3193 % Appropriate date and time representation (%c)
3194 % "%F %T"
3195 d_t_fmt "<%><F><SP><%><T>"
3196 %
3197 % Appropriate date representation (%x) "%F"
3198 d_fmt "<%><F>"
3199 %
3200 % Appropriate time representation (%X) "%T"
3201 t_fmt "<%><T>"
3202 t_fmt_ampm ""
3203 %
3204 END LC_TIME
3205
3206
3207
3208

```

#### 4.7 LC\_MESSAGES

The LC\_MESSAGES category shall define the format and values for affirmative and negative responses. The operands shall be strings or extended regular expressions to specify which response strings that should be considered matches; see ISO/IEC 9945-2:1993 clause 2.8.4 for a definition of extended regular expressions. The following keywords shall be defined:

- copy** Specify the name of an existing FDCC-set to be used as the source for the definition of this category. If this keyword is specified, no other keyword shall be specified.
- yesexpr** The operand shall consist of an extended regular expression that describes the acceptable affirmative response to a question expecting an affirmative or negative response.
- noexpr** The operand shall consist of an extended regular expression that describes the acceptable negative response to a question expecting an affirmative or negative response.

The "i18n" LC\_MESSAGES category is:

```

3225 LC_MESSAGES
3226 % This is the ISO/IEC 14652 "i18n" definition for
3227 % the LC_MESSAGES category.
3228 %
3229 yesexpr "<U005B><+><1><U005D>"
3230 noexpr "<U005B><-><0><U005D>"
3231 END LC_MESSAGES
3232
3233
3234

```

#### 4.8 LC\_PAPER

The LC\_PAPER category defines the default size of paper used for documents. The following keywords shall be defined:

- copy** Specify the name of an existing FDCC-set to be used as the source for the definition of this category. If this keyword is specified, no other keyword shall be specified.

3243 **height** Shall be used to specify the vertical dimension of the paper. The operand  
 3244 is an integer and the value is the height measured in millimetres.  
 3245 **width** Shall be used to specify the horizontal dimension of the paper. The  
 3246 operand is an integer and the value is the width measured in millimetres.  
 3247

3248 NOTE: If the height is greater than the width, it is called to be in portrait  
 3249 position, else it is called to be in landscape position.  
 3250

3251 The "i18n" LC\_PAPER category is:

```
3252 LC_PAPER
3253 % This is the ISO/IEC 14652 "i18n" definition for
3254 % the LC_PAPER category.
3255 %
3256 %
3257 height 297
3258 width 210
3259 END LC_PAPER
```

#### 3260 4.9 LC\_NAME

3261 The LC\_NAME category defines formats to be used in addressing a person, e.g. in a  
 3262 postal address or in a letter. The following keywords shall be defined:

3263 **copy** Specify the name of an existing FDCC-set to be used as the source for the  
 3264 definition of this category. If this keyword is specified, no other keyword  
 3265 shall be specified.  
 3266 **name\_fmt** Define the appropriate representation of a person's name and title. The  
 3267 operand shall consist of a string, and can contain any combination of  
 3268 characters and field descriptors. In addition, the string can contain escape  
 3269 sequences defined below.  
 3270 **name\_gen** The operand is a string defining a salutation valid for all persons,  
 3271 example: the Japanese "-sama" salutation in a letter.  
 3272 **name\_miss** The operand is a string defining a salutation valid for unmarried females.  
 3273 **name\_mr** The operand is a string defining a salutation valid for males.  
 3274 **name\_mrs** The operand is a string defining a salutation valid for married females.  
 3275 **name\_ms** The operand is a string defining a salutation valid for all females.  
 3276  
 3277  
 3278  
 3279

3280 NOTE: There are a number of variations for addressing a person among the cultures.  
 3281 Middle names are not used in many countries and even the family name is not used in  
 3282 some countries. The specification below should be regarded as a starting point for this  
 3283 problem.  
 3284

3285 The LC\_NAME category defines the interpretation of a number of escape sequences. The  
 3286 escape sequences are also available in the definitions with the following LC\_NAME  
 3287 keywords: "name\_fmt".  
 3288

3289 Escape sequences for the "name\_fmt" keyword:

```
3290
3291 %f      Family names.
3292 %F      Family names in uppercase.
3293 %g      First given name.
3294 %G      First given initial.
3295 %l      First given name with latin letters.
```

3296	%o	Other shorter name, eg. "Bill".
3297	%m	Middle names.
3298	%M	Middle initial.
3299	%p	Profession.
3300	%s	Salutation, such as "Doctor"
3301	%S	Abbreviated salutation, such as "Mr." or "Dr."
3302	%d	Salutation, using the FDCC-sets conventions, with 1 for the name_gen, 2 for name_mr, 3 for name_mrs, 4 for name_miss, 5 for name_ms.
3303		
3304	%t	If the preceding escape sequence resulted in an empty string, then the empty string, else a <space>.
3305		
3306		

Each escape sequence may have an <R> after the <%> to specify that the information is taken from a Romanized version string of the entity.

The "i18n" LC\_NAME category is:

```

3311 LC_NAME
3312 % This is the ISO/IEC 14652 "i18n" definition for
3313 % the LC_NAME category.
3314 %
3315 %
3316 name_fmt    "<%><p><%><t><%><g><%><t><%><m><%><t><%><f>"
3317 END LC_NAME

```

#### 4.10 LC\_ADDRESS

The LC\_ADDRESS category defines formats to be used in specifying a location like a person's living or office, for use in a postal address or in a letter, and other items of geographic nature. All keywords are optional. The following keywords shall be defined:

3325	<b>copy</b>	Specify the name of an existing FDCC-set to be used as the source for the definition of this category. If this keyword is specified, no other keyword shall be specified.
3326		
3327		
3328	<b>postal_fmt</b>	Define the appropriate representation of a postal address such as street and city. The proper formatting of a person's name and title is done with the "name_fmt" keyword of the LC_NAME category. The operand shall consist of a string, and can contain any combination of characters and field descriptors. In addition, the string can contain escape sequences defined below.
3329		
3330		
3331		
3332		
3333		
3334	<b>country_name</b>	The operand is a string with the name of the country in the language of the FDCC-set.
3335		
3336	<b>country_post</b>	The operand is a string with the abbreviation of the country, used for postal addresses, according to CEPT-MAILCODE.
3337		
3338	<b>country_ab2</b>	The operand is a string with the two-letter abbreviation of the country, according to ISO 3166.
3339		
3340	<b>country_ab3</b>	The operand is a string with the three-letter abbreviation of the country, according to ISO 3166.
3341		
3342	<b>country_num</b>	The operand is an integer with the three-digit number of the country, according to ISO 3166.
3343		
3344	<b>country_car</b>	The operand is a string with the abbreviation of the country, used for motor vehicles and traffic, according to the Genève convention 1949:68.
3345		
3346		
3347	<b>country_isbn</b>	The operand is a string with the abbreviation of the country, used for

3348		book numbering (ISBN), according to ISO 2108.
3349	<b>lang_name</b>	The operand is a string with the name of the language in the language of the FDCC-set.
3350		
3351	<b>lang_ab</b>	The operand is a string with the two-letter abbreviation of the language, according to ISO 639.
3352		
3353	<b>lang_term</b>	The operand is a string with the three-letter abbreviation of the language for terminology use, according to ISO 639-2.
3354		
3355	<b>lang_lib</b>	The operand is a string with the three-letter abbreviation of the language for library use, according to ISO 639-2. If not specified, the value of the "lang_term" keyword is taken.
3356		
3357		
3358		

The LC\_ADDRESS category defines the interpretation of a number of escape sequences.

The escape sequences are also available in the definitions with the following

LC\_ADDRESS keywords: "postal\_fmt".

Escape sequences for the "postal\_fmt" keyword:

3363		
3364		
3365	%a	C/O address.
3366	%f	Firm name.
3367	%d	department name.
3368	%b	Building name.
3369	%s	street or block (eg. Japanese) name.
3370	%h	house number or designation.
3371	%N	if any graphical characters have been specified then an end of line is made.
3372		
3373	%t	if the preceding escape sequence resulted in an empty string, then the empty string, else a <space>.
3374		
3375	%r	room number, door designation.
3376	%e	floor number.
3377	%C	country designation.
3378	%z	zip number, postal code.
3379	%T	town, city.
3380	%c	country.
3381		

Each escape sequence may have an <R> after the <%> to specify that the information is taken from a Romanized version string of the entity.

NOTE: There are a number of variations for specifying a location among the cultures. Some of the information, like the middle names, or even the family name, is not used in some cultures. The specification here should be regarded as a start point for this problem.

The "i18n" LC\_ADDRESS category is:

```
LC_ADDRESS
% This is the ISO/IEC 14652 "i18n" definition for
% the LC_ADDRESS category.
%
postal_fmt      "<%><a><%><N><%><f><%><N><%><d><%><N><%><b><%><N>/
<%><s><SP><%><h><SP><%><e><SP><%><r><%><N>/
<%><C><-><%><z><SP><%><T><%><N><%><c><%><N>"
END LC_ADDRESS
```

## 3402 4.11 LC\_TELEPHONE

3403

3404 The LC\_TELEPHONE category defines formats to be used with telephone services. All  
3405 keywords are optional. The following keywords shall be defined:

3406

3407 **copy** Specify the name of an existing FDCC-set to be used as the source  
3408 for the definition of this category. If this keyword is specified, no  
3409 other keyword shall be specified.

3410 **tel\_int\_fmt** Define the appropriate representation of a telephone number for  
3411 international use. The operand shall consist of a string, and can  
3412 contain any combination of characters and field descriptors. In  
3413 addition, the string can contain escape sequences defined below.

3414 **tel\_dom\_fmt** Define the appropriate representation of a telephone number for  
3415 domestic use. The operand shall consist of a string, and can contain  
3416 any combination of characters and field descriptors. In addition, the  
3417 string can contain escape sequences defined below.

3418 **int\_select** The operand is a string with the digits used to call international  
3419 telephone numbers.

3420 **int\_prefix** The operand is a string with the prefix used from other countries to  
3421 call the area

3422

3423 The LC\_TELEPHONE category defines the interpretation of a number of escape  
3424 sequences. The escape sequences are also available in the definitions with the following  
3425 LC\_TELEPHONE keywords: "tel\_int\_fmt" and "tel\_dom\_fmt".

3426

3427 %a area code without prefix (prefix is often <0>).

3428 %A area code including prefix (prefix is often <0>).

3429 %l local number.

3430 %c country code

3431 %C alternative carrier service code used for dialling abroad

3432

3433 The "i18n" LC\_TELEPHONE category is:

3434

```
3435 LC_TELEPHONE
```

```
3436 % This is the ISO/IEC 14652 "i18n" definition for
```

```
3437 % the LC_TELEPHONE category.
```

```
3438 %
```

```
3439 tel_int_fmt      "<+><%><c><SP><%><a><SP><%><l>"
```

```
3440 END LC_TELEPHONE
```

3441

3442

3443

## 3443 5. CHARMAP

3444

3445 A character set description may exist for each coded character set supported by an  
3446 application. This text is referred elsewhere in this standard as a charmap.

3447

3448 A conforming charmap to be used with a FDCC-set shall support the portable character set  
3449 specified in Table 1.

3450

3451 Conforming charmaps shall specify certain character and character set attributes, as  
3452 defined in 5.1.

3453

## 3454 5.1 Character Set Description Text

3455  
3456 The character set description text (charmap) describes the mapping between symbolic  
3457 character names and actual encoding of a coded character set. It is used to bind the  
3458 symbolic character names in a FDCC-set to an actual encoding, so an application can  
3459 process data in this encoding.

3460  
3461 The following declarations can precede the character definitions. Each shall consist of the  
3462 symbol shown in the following list, starting in column 1, including the surrounding  
3463 brackets, followed by one or more "blank"s, followed by the value to be assigned to the  
3464 symbol. If any of the declarations are included, they shall be specified in the order shown  
3465 in the following list:

3466		
3467	<b>&lt;code_set_name&gt;</b>	The name of the coded character set for which the character set description text is defined. The characters of the name shall be taken from the set of characters with visible glyphs defined in Table 1.
3468		
3469		
3470		
3471		
3472	<b>&lt;mb_cur_max&gt;</b>	The maximum number of bytes in a multibyte character. This shall default to 1.
3473		
3474		
3475	<b>&lt;mb_cur_min&gt;</b>	An unsigned positive integer value that shall define the minimum number of bytes in a character for the encoded character set. The value shall be less or equal to "mb_cur_max". If not specified, the minimum number shall be equal to "mb_cur_max".
3476		
3477		
3478		
3479		
3480		
3481	<b>&lt;escape_char&gt;</b>	The escape character used to indicate that the characters following shall be interpreted in a special way, as defined later in this subclause. This shall default to backslash (\). The character slash (/) is used in all the following text and examples, unless otherwise noted.
3482		
3483		
3484		
3485		
3486		
3487	<b>&lt;comment_char&gt;</b>	The character that when placed in column 1 of a charmap line, is used to indicate that the line shall be ignored. The default character shall be the number sign (#). The character percent-sign (%) is used in all the following text and examples, unless otherwise noted.
3488		
3489		
3490		
3491		
3492		
3493	<b>&lt;repertoiremap&gt;</b>	The name of the repertoiremap used to define the symbolic character names in the charmap. The characters of the name shall be taken from the set of characters with visible glyphs defined in Table 1.
3494		
3495		
3496		
3497		
3498	<b>&lt;escseq&gt;</b>	defines the escape sequences for ISO 2022 shifting for the coded character set defined by the charmap. The semicolon-separated operands are all strings with characters taken from the set of characters with visible glyphs defined in table 1. The first operand defines the g-set or c-set to be defined, and the following values are defined: c0, c1, g0, g1, g2, g3. The second
3499		
3500		
3501		
3502		
3503		

operand defines what range of characters in the charmap is affected, and the values defined are: c0, c1, g0, g1. The third operand is the escape sequence that is defined.

**<addset>** the name of the charmap to be added the current coded character set and to be selected by the escape sequences defined by <escseq> of the added charmap.

**<include>** include the encoding of another charmap in the current charmap. The semicolon-separated operands are all strings with characters taken from the set of characters with visible glyphs defined in table 1. The first operand defines the g-set or c-set to be defined in the current charmap, and the following values are defined: c0, c1, g0, g1, g2, g3. The second operand defines a range of characters in the referenced charmap, and the values defined are: c0, c1, g0, g1. The third operand is the name of the charmap to be included. The coded character sets are defined initially for the encoding, and therefore do not need escape sequences for identification. If two g0 sets are defined, the second is switched to using the SHIFT OUT control character, while the first is shifted to using the SHIFT IN control character.

The character set mapping definitions shall be all the lines immediately following an identifier line containing the string "CHARMAP" starting in column 1, and preceding a trailer line containing the string "END CHARMAP" starting in column 1. Empty lines and lines containing a <comment\_char> in the first column shall be ignored. Each noncomment line of the character set mapping definition (i.e., between the "CHARMAP" and "END CHARMAP" lines of the text) shall be in one of the following syntaxes.

"%s %s %s\n", <symbolic-name>,<encoding>,<comments>

"%s...%s %s %s\n", <symbolic-name>,<symbolic-name>,<encoding>,<comments>

"%s....%s %s %s\n", <symbolic-name>,<symbolic-name>,<encoding>,<comments>

"%s..%s %s %s\n", <symbolic-name>,<symbolic-name>,<encoding>,<comments>

In the first syntax, the line of the character set mapping definition shall start with the symbolic name, immediately preceded by a <less-than> character and immediately followed by a <greater-than> character. Symbolic names shall only contain characters from the set shown with a visible glyph in Table 1.

The same symbolic name may occur several times, with different values. The first value is the one used when generating an encoding, while the other values are accepted in decoding. Symbolic names may be included to identify values that can overlap with each other or with the values of the symbolic names shown in Table 1. It is possible to specify symbolic names for which no encoding exists in the encoded character set, by not specifying a value.



3554 In the second and third syntax (symbolic decimal ellipsis), the line in the character set  
3555 mapping defines a range of one or more symbolic names. The difference between the  
3556 second and the third syntax is the number of dots in the ellipsis: the second has 3 dots, the  
3557 third has 4 dots. In these forms the symbolic names shall consist of zero or more  
3558 nonnumeric characters from the set shown with visible glyphs in Table 1, followed by an  
3559 integer formed by one or more decimal digits. The characters preceding the integer shall  
3560 be identical in the two symbolic names, and the integer formed by the digits in the second  
3561 symbolic name shall be identical to or greater than the integer formed by the digits in the  
3562 first name. This shall be interpreted as a series of symbolic names formed from the  
3563 common part and each of the integers in decimal format between the first and the second  
3564 integer, inclusive, and with a length of the symbolic names generated that is equal to the  
3565 length of the first (and also the second) symbolic name. As an example,  
3566 <j0101>...<j0104> is interpreted as the symbolic names <j0101>, <j0102>, <j0103>, and  
3567 <j0104>, in that order.

3568

3569 Note: The rationale to allow both a 3-dot and a 4-dot symbol for symbolic decimal  
3570 ellipses is that in the POSIX standard the decimal symbolic ellipses was defined by a 3-  
3571 dot symbol for charmaps, while the 3-dot symbol was an absolute ellipses for POSIX  
3572 locales, and this International standard specifies a 4-dot symbol for the decimal  
3573 symbolic ellipses. The 3-dot symbolic decimal ellipses in charmaps is deprecated.

3574

3575 In the fourth syntax (symbolic hexadecimal ellipsis, with two dots), the line in the  
3576 character set mapping defines a range of one or more symbolic names. In this form the  
3577 symbolic names shall consist of zero or more nonnumeric characters from the set shown  
3578 with visible glyphs in Table 1, followed by an integer formed by one or more hexadecimal  
3579 digits, using uppercase letters only for the range "A" to "F". The characters preceding the  
3580 hexadecimal integer shall be identical in the two symbolic names, and the integer formed  
3581 by the hexadecimal digits in the second symbolic name shall be identical to or greater than  
3582 the integer formed by the hexadecimal digits in the first name. This shall be interpreted as  
3583 a series of symbolic names formed from the common part and each of the integers in  
3584 hexadecimal format using uppercase letters only between the first and the second integer,  
3585 inclusive, and with a length of the symbolic names generated that is equal to the length of  
3586 the first (and also the second) symbolic name. As an example, <U010E>..<U0111> is  
3587 interpreted as the symbolic names <U010E>, <U010F>, <U0110>, and <U0111>, in that  
3588 order.

3589

3590 The encoding part shall be expressed as one (for single-byte values) or more concatenated  
3591 decimal, octal or hexadecimal constants. Decimal constants shall be represented by two or  
3592 three decimal digits, preceded by the escape character and the lowercase letter "d"; for  
3593 example /d05, /d97, or /d143. Hexadecimal constants shall be represented by two  
3594 hexadecimal digits, preceded by the escape character and the lowercase letter "x"; for  
3595 example /x05, /x61, or /x8f. Octal constants shall be represented by two or three octal  
3596 digits, preceded by the escape character; for example /05, /141, or /217. In a charmap,  
3597 each constant should represent an 8 bit byte for portability reasons. Applications  
3598 supporting other byte sizes may allow constants to represent values larger than those that  
3599 can be represented in 8 bit bytes, and to allow additional digits in constants. When  
3600 constants are concatenated for multibyte character values, they may be of different types,  
3601 and interpreted in byte order from the first to the last with the least significant byte of the  
3602 multibyte character specified by the last byte. The manner in which these constants are  
3603 represented in the character stored in the system is application defined. Omitting bytes

3604 from a multibyte character produces undefined results.

3605

3606 In lines defining ranges of symbolic names, the encoded value is the value for the first  
3607 symbolic name in the range (the symbolic name preceding the ellipsis). Subsequent  
3608 symbolic names defined by the range shall have encoding values in increasing order. For  
3609 example the line

3610

3611 <j0101>....<j0104> /d129/d254

3612

3613 shall be interpreted as

3614

3615 <j0101> /d129/d254

3616

<j0102> /d129/d255

3617

<j0103> /d130/d000

3618

<j0104> /d130/d001

3619

3620 The comments parameter is optional.

3621

3622

3623 Example of using ISO 2022 techniques:

3624

3625 The following example defines two coded character sets, a 7-bit and a 14-bit. They are then merged into one  
3626 encoding. It is an example on how encodings used in Eastern Asia could be specified.

3627

3628 The 7-bit charmap

3629

3630 <escape\_char> /

3631

<comment\_char> %

3632

% The 7bit charmap defines both control and graphic characters

3633

<code\_set\_name> "eastern7bit"

3634

<escseq> "c0";"c0","/x21/x40"

3635

<escseq> "g0";"g0","/x28/x48"

3636

<escseq> "g1";"g0","/x29/x48"

3637

<escseq> "g2";"g0","/x2A/x48"

3638

<escseq> "g3";"g0","/x2B/x48"

3639

CHARMAP

3640

<tab> /x08

3641

<newline> /x0D

3642

<a> /x61

3643

% more character encodings to be defined here

3644

END CHARMAP

3645

3646

3647

3648 The 14-bit charmap

3649

3650 <escape\_char> /

3651

<comment\_char> %

3652

<code\_set\_name> "eastern14bit"

3653

<mb\_cur\_max> 2

3654

<esqseq> "g0";"g0","/x24/x40"

3655

<esqseq> "g1";"g0","/x24/x29/x40"

3656

<esqseq> "g2";"g0","/x24/x2A/x40"

3657

<esqseq> "g3";"g0","/x24/x2B/x40"

3658

CHARMAP

3659

<U0365> /d036/d055 % the character codes are only examples

```

3660      <U0744>      /d036/d056
3661      % more character encodings to be defined here
3662      END CHARMAP

```

The merged encoding

```

3667      <escape_char> /
3668      <comment_char> %
3669      <code_set_name>  "shift-eastern"
3670      <mb_cur_max>    2
3671      <mb_cur_min>    1
3672      <include>       "c0";"c0";"eastern7bit"
3673      <include>       "g0";"g0";"eastern7bit"
3674      <include>       "g1";"g0";"eastern14bit"
3675      % This defines the g0 values of "eastern14bit" (without the 8th
3676      % bit set) to be the g1 in this encoding (with the 8th bit set).
3677      %
3678      % So the bytes without the 8th bit set is from the "shift7bit"
3679      % coded character set, while bytes with the 8th bit set are from
3680      % the 14-bit set.

```

Another merged encoding using the same charmaps:

```

3684      <escape_char> /
3685      <comment_char> %
3686      <code_set_name>  "EUC-eastern"
3687      <mb_cur_max>    2
3688      <mb_cur_min>    1
3689      <include>       "c0";"c0";"eastern7bit"
3690      <include>       "g0";"g0";"eastern7bit"
3691      <include>       "g0";"g0";"eastern14bit"
3692      % As there are two "g0" sets defined, the first referenced is the
3693      % initial g0 set, while the second can be shifted to via the SHIFT OUT
3694      % control character. The first can then be shifted to by the SHIFT IN
3695      % control character.

```

## 6 REPertoireMAP

FDCC-set and Charmap sources may be specified in a coded character set independent way, using symbolic character names. The relation between the symbolic character names and characters may be specified via a Repertoiremap, which defines the repertoire of characters defined for a FDCC-set, and the symbolic character names and corresponding abstract character (by a reference to ISO/IEC 10646).

The repertoire mapping is defined by specifying the symbolic character name and the ISO/IEC 10646 code position in hexadecimal form (with a preceding 'U') and optionally the long ISO/IEC 10646 character name in the following syntax:

```
"%s %s %s\n",<symbolic-name>,<10646-short-identifier>,<comments>
```

The symbolic character name and the ISO/IEC 10646 short identifier are each surrounded by angle brackets <>, and the fields shall be separated by one or more spaces or tabs on a line. If a right angle bracket or an escape character is used within a symbolic name, it shall be preceded by the escape character. Characters not in ISO/IEC 10646 may be referenced by the symbolic character names <U80000000>..<U8FFFFFFF>.

3717 The escape character can be redefined from the default reverse solidus (\) with the first  
 3718 line of the Repertoiremap containing the string "escape\_char" followed by one or more  
 3719 spaces or tabs and then the escape character.

3720  
 3721 Several symbolic character names can refer to the same abstract character, and are then  
 3722 used as synonyms in FDCC-sets and charmaps. The set of <U0000>..
 3723 <U00000000>..
 3724 refers to the corresponding code points of ISO/IEC 10646 with the same short identifier.

3725  
 3726 The "i18nrep" repertoiremap is defined to accommodate prior art, such as defined in the  
 3727 ISO/IEC 9945-2:1993 standard annex G, and used by ISO and IEC member bodies in their  
 3728 national POSIX locale specifications, and as used in POSIX locales distributed by the  
 3729 ISO/IEC POSIX working group and X/Open. Many POSIX charmaps registered with  
 3730 ISO/IEC 15897 use these symbolic names. It also reflects use on the Internet, and many of  
 3731 the Internet registered charsets are specified using these symbolic names. The "i18nrep"  
 3732 repertoiremap thus facilitates reuse of both POSIX locale data and POSIX charmaps with  
 3733 data from this International Standard. The contents of the "i18nrep" repertoiremap is as  
 3734 follows:

3735	escape_char /	
3736	<NUL>	<U0000> NULL (NUL)
3737	<SOH>	<U0001> START OF HEADING (SOH)
3738	<STX>	<U0002> START OF TEXT (STX)
3739	<ETX>	<U0003> END OF TEXT (ETX)
3740	<EOT>	<U0004> END OF TRANSMISSION (EOT)
3741	<ENQ>	<U0005> ENQUIRY (ENQ)
3742	<ACK>	<U0006> ACKNOWLEDGE (ACK)
3743	<alert>	<U0007> BELL (BEL)
3744	<BEL>	<U0007> BELL (BEL)
3745	<backspace>	<U0008> BACKSPACE (BS)
3746	<tab>	<U0009> CHARACTER TABULATION (HT)
3747	<newline>	<U000A> LINE FEED (LF)
3748	<vertical-tab>	<U000B> LINE TABULATION (VT)
3749	<form-feed>	<U000C> FORM FEED (FF)
3750	<carriage-return>	<U000D> CARRIAGE RETURN (CR)
3751	<DLE>	<U0010> DATALINK ESCAPE (DLE)
3752	<DC1>	<U0011> DEVICE CONTROL ONE (DC1)
3753	<DC2>	<U0012> DEVICE CONTROL TWO (DC2)
3754	<DC3>	<U0013> DEVICE CONTROL THREE (DC3)
3755	<DC4>	<U0014> DEVICE CONTROL FOUR (DC4)
3756	<NAK>	<U0015> NEGATIVE ACKNOWLEDGE (NAK)
3757	<SYN>	<U0016> SYNCHRONOUS IDLE (SYN)
3758	<ETB>	<U0017> END OF TRANSMISSION BLOCK (ETB)
3759	<CAN>	<U0018> CANCEL (CAN)
3760	<SUB>	<U001A> SUBSTITUTE (SUB)
3761	<ESC>	<U001B> ESCAPE (ESC)
3762	<IS4>	<U001C> FILE SEPARATOR (IS4)
3763	<IS3>	<U001D> GROUP SEPARATOR (IS3)
3764	<intro>	<U001D> GROUP SEPARATOR (IS3)
3765	<IS2>	<U001E> RECORD SEPARATOR (IS2)
3766	<IS1>	<U001F> UNIT SEPARATOR (IS1)
3767	<DEL>	<U007F> DELETE (DEL)
3768	<space>	<U0020> SPACE
3769	<exclamation-mark>	<U0021> EXCLAMATION MARK
3770	<quotation-mark>	<U0022> QUOTATION MARK
3771	<number-sign>	<U0023> NUMBER SIGN
3772	<dollar-sign>	<U0024> DOLLAR SIGN
3773	<percent-sign>	<U0025> PERCENT SIGN
3774	<ampersand>	<U0026> AMPERSAND
3775	<apostrophe>	<U0027> APOSTROPHE
3776	<left-parenthesis>	<U0028> LEFT PARENTHESIS
3777	<right-parenthesis>	<U0029> RIGHT PARENTHESIS
3778	<asterisk>	<U002A> ASTERISK
3779	<plus-sign>	<U002B> PLUS SIGN
3780	<comma>	<U002C> COMMA
3781	<hyphen>	<U002D> HYPHEN-MINUS
3782	<hyphen-minus>	<U002D> HYPHEN-MINUS
3783	<period>	<U002E> FULL STOP
3784	<full-stop>	<U002E> FULL STOP
3785	<slash>	<U002F> SOLIDUS
3786	<solidus>	<U002F> SOLIDUS
3787	<zero>	<U0030> DIGIT ZERO
3788	<one>	<U0031> DIGIT ONE
3789	<two>	<U0032> DIGIT TWO
3790		

3791	<three>	<U0033>	DIGIT THREE
3792	<four>	<U0034>	DIGIT FOUR
3793	<five>	<U0035>	DIGIT FIVE
3794	<six>	<U0036>	DIGIT SIX
3795	<seven>	<U0037>	DIGIT SEVEN
3796	<eight>	<U0038>	DIGIT EIGHT
3797	<nine>	<U0039>	DIGIT NINE
3798	<colon>	<U003A>	COLON
3799	<semicolon>	<U003B>	SEMICOLON
3800	<less-than-sign>	<U003C>	LESS-THAN SIGN
3801	<equals-sign>	<U003D>	EQUALS SIGN
3802	<greater-than-sign>	<U003E>	GREATER-THAN SIGN
3803	<question-mark>	<U003F>	QUESTION MARK
3804	<commercial-at>	<U0040>	COMMERCIAL AT
3805	<left-square-bracket>	<U005B>	LEFT SQUARE BRACKET
3806	<backslash>	<U005C>	REVERSE SOLIDUS
3807	<reverse-solidus>	<U005C>	REVERSE SOLIDUS
3808	<right-square-bracket>	<U005D>	RIGHT SQUARE BRACKET
3809	<circumflex>	<U005E>	CIRCUMFLEX ACCENT
3810	<circumflex-accent>	<U005E>	CIRCUMFLEX ACCENT
3811	<underscore>	<U005F>	LOW LINE
3812	<low-line>	<U005F>	LOW LINE
3813	<grave-accent>	<U0060>	GRAVE ACCENT
3814	<left-brace>	<U007B>	LEFT CURLY BRACKET
3815	<left-curly-bracket>	<U007B>	LEFT CURLY BRACKET
3816	<vertical-line>	<U007C>	VERTICAL LINE
3817	<right-brace>	<U007D>	RIGHT CURLY BRACKET
3818	<right-curly-bracket>	<U007D>	RIGHT CURLY BRACKET
3819	<tilde>	<U007E>	TILDE
3820			
3821	<a8>	<U0252>	Weight indicating the position of the last a
3822	<b8>	<U0182>	Weight indicating the position of the last b
3823	<c8>	<U0255>	Weight indicating the position of the last c
3824	<d8>	<U018D>	Weight indicating the position of the last d
3825	<e8>	<U0264>	Weight indicating the position of the last e
3826	<f8>	<U0191>	Weight indicating the position of the last f
3827	<g8>	<U01A2>	Weight indicating the position of the last g
3828	<h8>	<U02BD>	Weight indicating the position of the last h
3829	<i8>	<U0196>	Weight indicating the position of the last i
3830	<j8>	<U0284>	Weight indicating the position of the last j
3831	<k8>	<U029E>	Weight indicating the position of the last k
3832	<l8>	<U028E>	Weight indicating the position of the last l
3833	<m8>	<U0271>	Weight indicating the position of the last m
3834	<n8>	<U014A>	Weight indicating the position of the last n
3835	<o8>	<U0277>	Weight indicating the position of the last o
3836	<p8>	<U0278>	Weight indicating the position of the last p
3837	<q8>	<U0138>	Weight indicating the position of the last q
3838	<r8>	<U02B6>	Weight indicating the position of the last r
3839	<s8>	<U0286>	Weight indicating the position of the last s
3840	<t8>	<U0287>	Weight indicating the position of the last t
3841	<u8>	<U01B1>	Weight indicating the position of the last u
3842	<v8>	<U028C>	Weight indicating the position of the last v
3843	<w8>	<U028D>	Weight indicating the position of the last w
3844	<x8>	<U216B>	Weight indicating the position of the last x
3845	<y8>	<U01B3>	Weight indicating the position of the last y
3846	<z8>	<U0293>	Weight indicating the position of the last z
3847			
3848	<NU>	<U0000>	NULL (NUL)
3849	<SH>	<U0001>	START OF HEADING (SOH)
3850	<SX>	<U0002>	START OF TEXT (STX)
3851	<EX>	<U0003>	END OF TEXT (ETX)
3852	<ET>	<U0004>	END OF TRANSMISSION (EOT)
3853	<EQ>	<U0005>	ENQUIRY (ENQ)
3854	<AK>	<U0006>	ACKNOWLEDGE (ACK)
3855	<BL>	<U0007>	BELL (BEL)
3856	<BS>	<U0008>	BACKSPACE (BS)
3857	<HT>	<U0009>	CHARACTER TABULATION (HT)
3858	<LF>	<U000A>	LINE FEED (LF)
3859	<VT>	<U000B>	LINE TABULATION (VT)
3860	<FF>	<U000C>	FORM FEED (FF)
3861	<CR>	<U000D>	CARRIAGE RETURN (CR)
3862	<SO>	<U000E>	SHIFT OUT (SO)
3863	<SI>	<U000F>	SHIFT IN (SI)
3864	<DL>	<U0010>	DATALINK ESCAPE (DLE)
3865	<D1>	<U0011>	DEVICE CONTROL ONE (DC1)
3866	<D2>	<U0012>	DEVICE CONTROL TWO (DC2)
3867	<D3>	<U0013>	DEVICE CONTROL THREE (DC3)
3868	<D4>	<U0014>	DEVICE CONTROL FOUR (DC4)
3869	<NK>	<U0015>	NEGATIVE ACKNOWLEDGE (NAK)
3870	<SY>	<U0016>	SYNCHRONOUS IDLE (SYN)
3871	<EB>	<U0017>	END OF TRANSMISSION BLOCK (ETB)
3872	<CN>	<U0018>	CANCEL (CAN)
3873	<EM>	<U0019>	END OF MEDIUM (EM)
3874	<SB>	<U001A>	SUBSTITUTE (SUB)
3875	<EC>	<U001B>	ESCAPE (ESC)
3876	<FS>	<U001C>	FILE SEPARATOR (IS4)
3877	<GS>	<U001D>	GROUP SEPARATOR (IS3)
3878	<RS>	<U001E>	RECORD SEPARATOR (IS2)
3879	<US>	<U001F>	UNIT SEPARATOR (IS1)

3880	<DT>	<U007F>	DELETE (DEL)
3881	<PA>	<U0080>	PADDING CHARACTER (PAD)
3882	<HO>	<U0081>	HIGH OCTET PRESET (HOP)
3883	<BH>	<U0082>	BREAK PERMITTED HERE (BPH)
3884	<NH>	<U0083>	NO BREAK HERE (NBH)
3885	<IN>	<U0084>	INDEX (IND)
3886	<NL>	<U0085>	NEXT LINE (NEL)
3887	<SA>	<U0086>	START OF SELECTED AREA (SSA)
3888	<ES>	<U0087>	END OF SELECTED AREA (ESA)
3889	<HS>	<U0088>	CHARACTER TABULATION SET (HTS)
3890	<HJ>	<U0089>	CHARACTER TABULATION WITH JUSTIFICATION (HTJ)
3891	<VS>	<U008A>	LINE TABULATION SET (VTS)
3892	<PD>	<U008B>	PARTIAL LINE FORWARD (PLD)
3893	<PU>	<U008C>	PARTIAL LINE BACKWARD (PLU)
3894	<RI>	<U008D>	REVERSE LINE FEED (RI)
3895	<S2>	<U008E>	SINGLE-SHIFT TWO (SS2)
3896	<S3>	<U008F>	SINGLE-SHIFT THREE (SS3)
3897	<DC>	<U0090>	DEVICE CONTROL STRING (DCS)
3898	<P1>	<U0091>	PRIVATE USE ONE (PU1)
3899	<P2>	<U0092>	PRIVATE USE TWO (PU2)
3900	<TS>	<U0093>	SET TRANSMIT STATE (STS)
3901	<CC>	<U0094>	CANCEL CHARACTER (CCH)
3902	<MW>	<U0095>	MESSAGE WAITING (MW)
3903	<SG>	<U0096>	START OF GUARDED AREA (SPA)
3904	<EG>	<U0097>	END OF GUARDED AREA (EPA)
3905	<SS>	<U0098>	START OF STRING (SOS)
3906	<GC>	<U0099>	SINGLE GRAPHIC CHARACTER INTRODUCER (SGCI)
3907	<SC>	<U009A>	SINGLE CHARACTER INTRODUCER (SCI)
3908	<CI>	<U009B>	CONTROL SEQUENCE INTRODUCER (CSI)
3909	<ST>	<U009C>	STRING TERMINATOR (ST)
3910	<OC>	<U009D>	OPERATING SYSTEM COMMAND (OSC)
3911	<PM>	<U009E>	PRIVACY MESSAGE (PM)
3912	<AC>	<U009F>	APPLICATION PROGRAM COMMAND (APC)
3913	<SP>	<U0020>	SPACE
3914	<!>	<U0021>	EXCLAMATION MARK
3915	<">	<U0022>	QUOTATION MARK
3916	<#>	<U0023>	NUMBER SIGN
3917	<\$>	<U0024>	DOLLAR SIGN
3918	<%>	<U0025>	PERCENT SIGN
3919	<&>	<U0026>	AMPERSAND
3920	<'>	<U0027>	APOSTROPHE
3921	<(>	<U0028>	LEFT PARENTHESIS
3922	<)>	<U0029>	RIGHT PARENTHESIS
3923	<*>	<U002A>	ASTERISK
3924	<+>	<U002B>	PLUS SIGN
3925	<, >	<U002C>	COMMA
3926	<- >	<U002D>	HYPHEN-MINUS
3927	<.>	<U002E>	FULL STOP
3928	</ >	<U002F>	SOLIDUS
3929	<0>	<U0030>	DIGIT ZERO
3930	<1>	<U0031>	DIGIT ONE
3931	<2>	<U0032>	DIGIT TWO
3932	<3>	<U0033>	DIGIT THREE
3933	<4>	<U0034>	DIGIT FOUR
3934	<5>	<U0035>	DIGIT FIVE
3935	<6>	<U0036>	DIGIT SIX
3936	<7>	<U0037>	DIGIT SEVEN
3937	<8>	<U0038>	DIGIT EIGHT
3938	<9>	<U0039>	DIGIT NINE
3939	<:>	<U003A>	COLON
3940	<:>	<U003B>	SEMICOLON
3941	<<>	<U003C>	LESS-THAN SIGN
3942	<=>	<U003D>	EQUALS SIGN
3943	</>>	<U003E>	GREATER-THAN SIGN
3944	<?>	<U003F>	QUESTION MARK
3945	<@>	<U0040>	COMMERCIAL AT
3946	<A>	<U0041>	LATIN CAPITAL LETTER A
3947	<B>	<U0042>	LATIN CAPITAL LETTER B
3948	<C>	<U0043>	LATIN CAPITAL LETTER C
3949	<D>	<U0044>	LATIN CAPITAL LETTER D
3950	<E>	<U0045>	LATIN CAPITAL LETTER E
3951	<F>	<U0046>	LATIN CAPITAL LETTER F
3952	<G>	<U0047>	LATIN CAPITAL LETTER G
3953	<H>	<U0048>	LATIN CAPITAL LETTER H
3954	<I>	<U0049>	LATIN CAPITAL LETTER I
3955	<J>	<U004A>	LATIN CAPITAL LETTER J
3956	<K>	<U004B>	LATIN CAPITAL LETTER K
3957	<L>	<U004C>	LATIN CAPITAL LETTER L
3958	<M>	<U004D>	LATIN CAPITAL LETTER M
3959	<N>	<U004E>	LATIN CAPITAL LETTER N
3960	<O>	<U004F>	LATIN CAPITAL LETTER O
3961	<P>	<U0050>	LATIN CAPITAL LETTER P
3962	<Q>	<U0051>	LATIN CAPITAL LETTER Q
3963	<R>	<U0052>	LATIN CAPITAL LETTER R
3964	<S>	<U0053>	LATIN CAPITAL LETTER S
3965	<T>	<U0054>	LATIN CAPITAL LETTER T
3966	<U>	<U0055>	LATIN CAPITAL LETTER U

3967	<V>	<U0056>	LATIN CAPITAL LETTER V
3968	<W>	<U0057>	LATIN CAPITAL LETTER W
3969	<X>	<U0058>	LATIN CAPITAL LETTER X
3970	<Y>	<U0059>	LATIN CAPITAL LETTER Y
3971	<Z>	<U005A>	LATIN CAPITAL LETTER Z
3972	<<(>	<U005B>	LEFT SQUARE BRACKET
3973	</////>	<U005C>	REVERSE SOLIDUS
3974	<)/>>	<U005D>	RIGHT SQUARE BRACKET
3975	<' />>	<U005E>	CIRCUMFLEX ACCENT
3976	<_>	<U005F>	LOW LINE
3977	<'!>	<U0060>	GRAVE ACCENT
3978	<a>	<U0061>	LATIN SMALL LETTER A
3979	<b>	<U0062>	LATIN SMALL LETTER B
3980	<c>	<U0063>	LATIN SMALL LETTER C
3981	<d>	<U0064>	LATIN SMALL LETTER D
3982	<e>	<U0065>	LATIN SMALL LETTER E
3983	<f>	<U0066>	LATIN SMALL LETTER F
3984	<g>	<U0067>	LATIN SMALL LETTER G
3985	<h>	<U0068>	LATIN SMALL LETTER H
3986	<i>	<U0069>	LATIN SMALL LETTER I
3987	<j>	<U006A>	LATIN SMALL LETTER J
3988	<k>	<U006B>	LATIN SMALL LETTER K
3989	<l>	<U006C>	LATIN SMALL LETTER L
3990	<m>	<U006D>	LATIN SMALL LETTER M
3991	<n>	<U006E>	LATIN SMALL LETTER N
3992	<o>	<U006F>	LATIN SMALL LETTER O
3993	<p>	<U0070>	LATIN SMALL LETTER P
3994	<q>	<U0071>	LATIN SMALL LETTER Q
3995	<r>	<U0072>	LATIN SMALL LETTER R
3996	<s>	<U0073>	LATIN SMALL LETTER S
3997	<t>	<U0074>	LATIN SMALL LETTER T
3998	<u>	<U0075>	LATIN SMALL LETTER U
3999	<v>	<U0076>	LATIN SMALL LETTER V
4000	<w>	<U0077>	LATIN SMALL LETTER W
4001	<x>	<U0078>	LATIN SMALL LETTER X
4002	<y>	<U0079>	LATIN SMALL LETTER Y
4003	<z>	<U007A>	LATIN SMALL LETTER Z
4004	<(!>	<U007B>	LEFT CURLY BRACKET
4005	<!!>	<U007C>	VERTICAL LINE
4006	<!)>	<U007D>	RIGHT CURLY BRACKET
4007	<'?>	<U007E>	TILDE
4008	<NS>	<U00A0>	NO-BREAK SPACE
4009	<!I>	<U00A1>	INVERTED EXCLAMATION MARK
4010	<Ct>	<U00A2>	CENT SIGN
4011	<Pd>	<U00A3>	POUND SIGN
4012	<Cu>	<U00A4>	CURRENCY SIGN
4013	<Ye>	<U00A5>	YEN SIGN
4014	<BB>	<U00A6>	BROKEN BAR
4015	<SE>	<U00A7>	SECTION SIGN
4016	<' :>	<U00A8>	DIAERESIS
4017	<Co>	<U00A9>	COPYRIGHT SIGN
4018	<-a>	<U00AA>	FEMININE ORDINAL INDICATOR
4019	<<<>	<U00AB>	LEFT-POINTING DOUBLE ANGLE QUOTATION MARK
4020	<NO>	<U00AC>	NOT SIGN
4021	<-->	<U00AD>	SOFT HYPHEN
4022	<Rg>	<U00AE>	REGISTERED SIGN
4023	<'m>	<U00AF>	MACRON
4024	<DG>	<U00B0>	DEGREE SIGN
4025	<+>	<U00B1>	PLUS-MINUS SIGN
4026	<2S>	<U00B2>	SUPERSCRRIPT TWO
4027	<3S>	<U00B3>	SUPERSCRRIPT THREE
4028	<' ' >	<U00B4>	ACUTE ACCENT
4029	<My>	<U00B5>	MICRO SIGN
4030	<PI>	<U00B6>	PILCROW SIGN
4031	<.M>	<U00B7>	MIDDLE DOT
4032	<' ,>	<U00B8>	CEDILLA
4033	<1S>	<U00B9>	SUPERSCRRIPT ONE
4034	<-o>	<U00BA>	MASCULINE ORDINAL INDICATOR
4035	</>/>>	<U00BB>	RIGHT-POINTING DOUBLE ANGLE QUOTATION MARK
4036	<14>	<U00BC>	VULGAR FRACTION ONE QUARTER
4037	<12>	<U00BD>	VULGAR FRACTION ONE HALF
4038	<34>	<U00BE>	VULGAR FRACTION THREE QUARTERS
4039	<?I>	<U00BF>	INVERTED QUESTION MARK
4040	<A!>	<U00C0>	LATIN CAPITAL LETTER A WITH GRAVE
4041	<A'>	<U00C1>	LATIN CAPITAL LETTER A WITH ACUTE
4042	<A/ >>	<U00C2>	LATIN CAPITAL LETTER A WITH CIRCUMFLEX
4043	<A?>	<U00C3>	LATIN CAPITAL LETTER A WITH TILDE
4044	<A :>	<U00C4>	LATIN CAPITAL LETTER A WITH DIAERESIS
4045	<AA>	<U00C5>	LATIN CAPITAL LETTER A WITH RING ABOVE
4046	<AE>	<U00C6>	LATIN CAPITAL LETTER AE (ash)
4047	<C,>	<U00C7>	LATIN CAPITAL LETTER C WITH CEDILLA
4048	<E!>	<U00C8>	LATIN CAPITAL LETTER E WITH GRAVE
4049	<E'>	<U00C9>	LATIN CAPITAL LETTER E WITH ACUTE
4050	<E/ >>	<U00CA>	LATIN CAPITAL LETTER E WITH CIRCUMFLEX
4051	<E :>	<U00CB>	LATIN CAPITAL LETTER E WITH DIAERESIS
4052	<I!>	<U00CC>	LATIN CAPITAL LETTER I WITH GRAVE
4053	<I'>	<U00CD>	LATIN CAPITAL LETTER I WITH ACUTE
4054	<I/ >>	<U00CE>	LATIN CAPITAL LETTER I WITH CIRCUMFLEX
4055	<I :>	<U00CF>	LATIN CAPITAL LETTER I WITH DIAERESIS

4056	<D->	<U00D0>	LATIN CAPITAL LETTER ETH (Icelandic)
4057	<N?>	<U00D1>	LATIN CAPITAL LETTER N WITH TILDE
4058	<O!>	<U00D2>	LATIN CAPITAL LETTER O WITH GRAVE
4059	<O'>	<U00D3>	LATIN CAPITAL LETTER O WITH ACUTE
4060	<O/>>	<U00D4>	LATIN CAPITAL LETTER O WITH CIRCUMFLEX
4061	<O?>	<U00D5>	LATIN CAPITAL LETTER O WITH TILDE
4062	<O:>	<U00D6>	LATIN CAPITAL LETTER O WITH DIAERESIS
4063	<*X>	<U00D7>	MULTIPLICATION SIGN
4064	<O//>	<U00D8>	LATIN CAPITAL LETTER O WITH STROKE
4065	<U!>	<U00D9>	LATIN CAPITAL LETTER U WITH GRAVE
4066	<U'>	<U00DA>	LATIN CAPITAL LETTER U WITH ACUTE
4067	<U/>>	<U00DB>	LATIN CAPITAL LETTER U WITH CIRCUMFLEX
4068	<U:>	<U00DC>	LATIN CAPITAL LETTER U WITH DIAERESIS
4069	<Y'>	<U00DD>	LATIN CAPITAL LETTER Y WITH ACUTE
4070	<TH>	<U00DE>	LATIN CAPITAL LETTER THORN (Icelandic)
4071	<ss>	<U00DF>	LATIN SMALL LETTER SHARP S (German)
4072	<a!>	<U00E0>	LATIN SMALL LETTER A WITH GRAVE
4073	<a'>	<U00E1>	LATIN SMALL LETTER A WITH ACUTE
4074	<a/>>	<U00E2>	LATIN SMALL LETTER A WITH CIRCUMFLEX
4075	<a?>	<U00E3>	LATIN SMALL LETTER A WITH TILDE
4076	<a:>	<U00E4>	LATIN SMALL LETTER A WITH DIAERESIS
4077	<aa>	<U00E5>	LATIN SMALL LETTER A WITH RING ABOVE
4078	<ae>	<U00E6>	LATIN SMALL LETTER AE (ash)
4079	<c,>	<U00E7>	LATIN SMALL LETTER C WITH CEDILLA
4080	<e!>	<U00E8>	LATIN SMALL LETTER E WITH GRAVE
4081	<e'>	<U00E9>	LATIN SMALL LETTER E WITH ACUTE
4082	<e/>>	<U00EA>	LATIN SMALL LETTER E WITH CIRCUMFLEX
4083	<e:>	<U00EB>	LATIN SMALL LETTER E WITH DIAERESIS
4084	<i!>	<U00EC>	LATIN SMALL LETTER I WITH GRAVE
4085	<i'>	<U00ED>	LATIN SMALL LETTER I WITH ACUTE
4086	<i/>>	<U00EE>	LATIN SMALL LETTER I WITH CIRCUMFLEX
4087	<i:>	<U00EF>	LATIN SMALL LETTER I WITH DIAERESIS
4088	<d->	<U00F0>	LATIN SMALL LETTER ETH (Icelandic)
4089	<n?>	<U00F1>	LATIN SMALL LETTER N WITH TILDE
4090	<o!>	<U00F2>	LATIN SMALL LETTER O WITH GRAVE
4091	<o'>	<U00F3>	LATIN SMALL LETTER O WITH ACUTE
4092	<o/>>	<U00F4>	LATIN SMALL LETTER O WITH CIRCUMFLEX
4093	<o?>	<U00F5>	LATIN SMALL LETTER O WITH TILDE
4094	<o:>	<U00F6>	LATIN SMALL LETTER O WITH DIAERESIS
4095	<-:>	<U00F7>	DIVISION SIGN
4096	<o//>	<U00F8>	LATIN SMALL LETTER O WITH STROKE
4097	<u!>	<U00F9>	LATIN SMALL LETTER U WITH GRAVE
4098	<u'>	<U00FA>	LATIN SMALL LETTER U WITH ACUTE
4099	<u/>>	<U00FB>	LATIN SMALL LETTER U WITH CIRCUMFLEX
4100	<u:>	<U00FC>	LATIN SMALL LETTER U WITH DIAERESIS
4101	<y'>	<U00FD>	LATIN SMALL LETTER Y WITH ACUTE
4102	<th>	<U00FE>	LATIN SMALL LETTER THORN (Icelandic)
4103	<y:>	<U00FF>	LATIN SMALL LETTER Y WITH DIAERESIS
4104	<A->	<U0100>	LATIN CAPITAL LETTER A WITH MACRON
4105	<a->	<U0101>	LATIN SMALL LETTER A WITH MACRON
4106	<A(>	<U0102>	LATIN CAPITAL LETTER A WITH BREVE
4107	<a(>	<U0103>	LATIN SMALL LETTER A WITH BREVE
4108	<A; >	<U0104>	LATIN CAPITAL LETTER A WITH OGONEK
4109	<a; >	<U0105>	LATIN SMALL LETTER A WITH OGONEK
4110	<C'>	<U0106>	LATIN CAPITAL LETTER C WITH ACUTE
4111	<c'>	<U0107>	LATIN SMALL LETTER C WITH ACUTE
4112	<C/>>	<U0108>	LATIN CAPITAL LETTER C WITH CIRCUMFLEX
4113	<c/>>	<U0109>	LATIN SMALL LETTER C WITH CIRCUMFLEX
4114	<C.>	<U010A>	LATIN CAPITAL LETTER C WITH DOT ABOVE
4115	<c.>	<U010B>	LATIN SMALL LETTER C WITH DOT ABOVE
4116	<C<>	<U010C>	LATIN CAPITAL LETTER C WITH CARON
4117	<c<>	<U010D>	LATIN SMALL LETTER C WITH CARON
4118	<D<>	<U010E>	LATIN CAPITAL LETTER D WITH CARON
4119	<d<>	<U010F>	LATIN SMALL LETTER D WITH CARON
4120	<D//>	<U0110>	LATIN CAPITAL LETTER D WITH STROKE
4121	<d//>	<U0111>	LATIN SMALL LETTER D WITH STROKE
4122	<E->	<U0112>	LATIN CAPITAL LETTER E WITH MACRON
4123	<e->	<U0113>	LATIN SMALL LETTER E WITH MACRON
4124	<E(>	<U0114>	LATIN CAPITAL LETTER E WITH BREVE
4125	<e(>	<U0115>	LATIN SMALL LETTER E WITH BREVE
4126	<E.>	<U0116>	LATIN CAPITAL LETTER E WITH DOT ABOVE
4127	<e.>	<U0117>	LATIN SMALL LETTER E WITH DOT ABOVE
4128	<E; >	<U0118>	LATIN CAPITAL LETTER E WITH OGONEK
4129	<e; >	<U0119>	LATIN SMALL LETTER E WITH OGONEK
4130	<E<>	<U011A>	LATIN CAPITAL LETTER E WITH CARON
4131	<e<>	<U011B>	LATIN SMALL LETTER E WITH CARON
4132	<G/>>	<U011C>	LATIN CAPITAL LETTER G WITH CIRCUMFLEX
4133	<g/>>	<U011D>	LATIN SMALL LETTER G WITH CIRCUMFLEX
4134	<G(>	<U011E>	LATIN CAPITAL LETTER G WITH BREVE
4135	<g(>	<U011F>	LATIN SMALL LETTER G WITH BREVE
4136	<G.>	<U0120>	LATIN CAPITAL LETTER G WITH DOT ABOVE
4137	<g.>	<U0121>	LATIN SMALL LETTER G WITH DOT ABOVE
4138	<G;>	<U0122>	LATIN CAPITAL LETTER G WITH CEDILLA
4139	<g;>	<U0123>	LATIN SMALL LETTER G WITH CEDILLA
4140	<H/>>	<U0124>	LATIN CAPITAL LETTER H WITH CIRCUMFLEX
4141	<h/>>	<U0125>	LATIN SMALL LETTER H WITH CIRCUMFLEX
4142	<H//>	<U0126>	LATIN CAPITAL LETTER H WITH STROKE



4143	<h//>	<U0127>	LATIN SMALL LETTER H WITH STROKE
4144	<I?>	<U0128>	LATIN CAPITAL LETTER I WITH TILDE
4145	<i?>	<U0129>	LATIN SMALL LETTER I WITH TILDE
4146	<I->	<U012A>	LATIN CAPITAL LETTER I WITH MACRON
4147	<i->	<U012B>	LATIN SMALL LETTER I WITH MACRON
4148	<I(>	<U012C>	LATIN CAPITAL LETTER I WITH BREVE
4149	<i(>	<U012D>	LATIN SMALL LETTER I WITH BREVE
4150	<I; >	<U012E>	LATIN CAPITAL LETTER I WITH OGONEK
4151	<i; >	<U012F>	LATIN SMALL LETTER I WITH OGONEK
4152	<I. >	<U0130>	LATIN CAPITAL LETTER I WITH DOT ABOVE
4153	<i. >	<U0131>	LATIN SMALL LETTER DOTLESS I
4154	<IJ>	<U0132>	LATIN CAPITAL LIGATURE IJ
4155	<ij>	<U0133>	LATIN SMALL LIGATURE IJ
4156	<J/>>	<U0134>	LATIN CAPITAL LETTER J WITH CIRCUMFLEX
4157	<j/>>	<U0135>	LATIN SMALL LETTER J WITH CIRCUMFLEX
4158	<K, >	<U0136>	LATIN CAPITAL LETTER K WITH CEDILLA
4159	<k, >	<U0137>	LATIN SMALL LETTER K WITH CEDILLA
4160	<kk>	<U0138>	LATIN SMALL LETTER KRA (Greenlandic)
4161	<L' >	<U0139>	LATIN CAPITAL LETTER L WITH ACUTE
4162	<l' >	<U013A>	LATIN SMALL LETTER L WITH ACUTE
4163	<L, >	<U013B>	LATIN CAPITAL LETTER L WITH CEDILLA
4164	<l, >	<U013C>	LATIN SMALL LETTER L WITH CEDILLA
4165	<L< >	<U013D>	LATIN CAPITAL LETTER L WITH CARON
4166	<l< >	<U013E>	LATIN SMALL LETTER L WITH CARON
4167	<L. >	<U013F>	LATIN CAPITAL LETTER L WITH MIDDLE DOT
4168	<l. >	<U0140>	LATIN SMALL LETTER L WITH MIDDLE DOT
4169	<L//>	<U0141>	LATIN CAPITAL LETTER L WITH STROKE
4170	<l//>	<U0142>	LATIN SMALL LETTER L WITH STROKE
4171	<N' >	<U0143>	LATIN CAPITAL LETTER N WITH ACUTE
4172	<n' >	<U0144>	LATIN SMALL LETTER N WITH ACUTE
4173	<N, >	<U0145>	LATIN CAPITAL LETTER N WITH CEDILLA
4174	<n, >	<U0146>	LATIN SMALL LETTER N WITH CEDILLA
4175	<N< >	<U0147>	LATIN CAPITAL LETTER N WITH CARON
4176	<n< >	<U0148>	LATIN SMALL LETTER N WITH CARON
4177	<'n>	<U0149>	LATIN SMALL LETTER N PRECEDED BY APOSTROPHE
4178	<NG>	<U014A>	LATIN CAPITAL LETTER ENG (Sami)
4179	<ng>	<U014B>	LATIN SMALL LETTER ENG (Sami)
4180	<O->	<U014C>	LATIN CAPITAL LETTER O WITH MACRON
4181	<o->	<U014D>	LATIN SMALL LETTER O WITH MACRON
4182	<O(>	<U014E>	LATIN CAPITAL LETTER O WITH BREVE
4183	<o(>	<U014F>	LATIN SMALL LETTER O WITH BREVE
4184	<O" >	<U0150>	LATIN CAPITAL LETTER O WITH DOUBLE ACUTE
4185	<o" >	<U0151>	LATIN SMALL LETTER O WITH DOUBLE ACUTE
4186	<OE>	<U0152>	LATIN CAPITAL LIGATURE OE
4187	<oe>	<U0153>	LATIN SMALL LIGATURE OE
4188	<R' >	<U0154>	LATIN CAPITAL LETTER R WITH ACUTE
4189	<r' >	<U0155>	LATIN SMALL LETTER R WITH ACUTE
4190	<R, >	<U0156>	LATIN CAPITAL LETTER R WITH CEDILLA
4191	<r, >	<U0157>	LATIN SMALL LETTER R WITH CEDILLA
4192	<R< >	<U0158>	LATIN CAPITAL LETTER R WITH CARON
4193	<r< >	<U0159>	LATIN SMALL LETTER R WITH CARON
4194	<S' >	<U015A>	LATIN CAPITAL LETTER S WITH ACUTE
4195	<s' >	<U015B>	LATIN SMALL LETTER S WITH ACUTE
4196	<S/>>	<U015C>	LATIN CAPITAL LETTER S WITH CIRCUMFLEX
4197	<s/>>	<U015D>	LATIN SMALL LETTER S WITH CIRCUMFLEX
4198	<S, >	<U015E>	LATIN CAPITAL LETTER S WITH CEDILLA
4199	<s, >	<U015F>	LATIN SMALL LETTER S WITH CEDILLA
4200	<S< >	<U0160>	LATIN CAPITAL LETTER S WITH CARON
4201	<s< >	<U0161>	LATIN SMALL LETTER S WITH CARON
4202	<T, >	<U0162>	LATIN CAPITAL LETTER T WITH CEDILLA
4203	<t, >	<U0163>	LATIN SMALL LETTER T WITH CEDILLA
4204	<T< >	<U0164>	LATIN CAPITAL LETTER T WITH CARON
4205	<t< >	<U0165>	LATIN SMALL LETTER T WITH CARON
4206	<T//>	<U0166>	LATIN CAPITAL LETTER T WITH STROKE
4207	<t//>	<U0167>	LATIN SMALL LETTER T WITH STROKE
4208	<U?>	<U0168>	LATIN CAPITAL LETTER U WITH TILDE
4209	<u?>	<U0169>	LATIN SMALL LETTER U WITH TILDE
4210	<U->	<U016A>	LATIN CAPITAL LETTER U WITH MACRON
4211	<u->	<U016B>	LATIN SMALL LETTER U WITH MACRON
4212	<U(>	<U016C>	LATIN CAPITAL LETTER U WITH BREVE
4213	<u(>	<U016D>	LATIN SMALL LETTER U WITH BREVE
4214	<U0>	<U016E>	LATIN CAPITAL LETTER U WITH RING ABOVE
4215	<u0>	<U016F>	LATIN SMALL LETTER U WITH RING ABOVE
4216	<U" >	<U0170>	LATIN CAPITAL LETTER U WITH DOUBLE ACUTE
4217	<u" >	<U0171>	LATIN SMALL LETTER U WITH DOUBLE ACUTE
4218	<U; >	<U0172>	LATIN CAPITAL LETTER U WITH OGONEK
4219	<u; >	<U0173>	LATIN SMALL LETTER U WITH OGONEK
4220	<W/>>	<U0174>	LATIN CAPITAL LETTER W WITH CIRCUMFLEX
4221	<w/>>	<U0175>	LATIN SMALL LETTER W WITH CIRCUMFLEX
4222	<Y/>>	<U0176>	LATIN CAPITAL LETTER Y WITH CIRCUMFLEX
4223	<y/>>	<U0177>	LATIN SMALL LETTER Y WITH CIRCUMFLEX
4224	<Y: >	<U0178>	LATIN CAPITAL LETTER Y WITH DIAERESIS
4225	<z' >	<U0179>	LATIN CAPITAL LETTER Z WITH ACUTE
4226	<z' >	<U017A>	LATIN SMALL LETTER Z WITH ACUTE
4227	<Z. >	<U017B>	LATIN CAPITAL LETTER Z WITH DOT ABOVE
4228	<z. >	<U017C>	LATIN SMALL LETTER Z WITH DOT ABOVE
4229	<Z< >	<U017D>	LATIN CAPITAL LETTER Z WITH CARON
4230	<z< >	<U017E>	LATIN SMALL LETTER Z WITH CARON
4231	<sl>	<U017F>	LATIN SMALL LETTER LONG S

4232	<b//>	<U0180>	LATIN SMALL LETTER B WITH STROKE
4233	<B2>	<U0181>	LATIN CAPITAL LETTER B WITH HOOK
4234	<C2>	<U0187>	LATIN CAPITAL LETTER C WITH HOOK
4235	<c2>	<U0188>	LATIN SMALL LETTER C WITH HOOK
4236	<F2>	<U0191>	LATIN CAPITAL LETTER F WITH HOOK
4237	<f2>	<U0192>	LATIN SMALL LETTER F WITH HOOK
4238	<K2>	<U0198>	LATIN CAPITAL LETTER K WITH HOOK
4239	<k2>	<U0199>	LATIN SMALL LETTER K WITH HOOK
4240	<O9>	<U01A0>	LATIN CAPITAL LETTER O WITH HORN
4241	<o9>	<U01A1>	LATIN SMALL LETTER O WITH HORN
4242	<OI>	<U01A2>	LATIN CAPITAL LETTER OI
4243	<oi>	<U01A3>	LATIN SMALL LETTER OI
4244	<yr>	<U01A6>	LATIN LETTER YR
4245	<U9>	<U01AF>	LATIN CAPITAL LETTER U WITH HORN
4246	<u9>	<U01B0>	LATIN SMALL LETTER U WITH HORN
4247	<Z//>	<U01B5>	LATIN CAPITAL LETTER Z WITH STROKE
4248	<z//>	<U01B6>	LATIN SMALL LETTER Z WITH STROKE
4249	<ED>	<U01B7>	LATIN CAPITAL LETTER EZH
4250	<DZ<>	<U01C4>	LATIN CAPITAL LETTER DZ WITH CARON
4251	<Dz<>	<U01C5>	LATIN CAPITAL LETTER D WITH SMALL LETTER Z WITH CARON
4252	<dz<>	<U01C6>	LATIN SMALL LETTER DZ WITH CARON
4253	<LJ3>	<U01C7>	LATIN CAPITAL LETTER LJ
4254	<Lj3>	<U01C8>	LATIN CAPITAL LETTER L WITH SMALL LETTER J
4255	<lj3>	<U01C9>	LATIN SMALL LETTER LJ
4256	<NJ3>	<U01CA>	LATIN CAPITAL LETTER NJ
4257	<Nj3>	<U01CB>	LATIN CAPITAL LETTER N WITH SMALL LETTER J
4258	<nj3>	<U01CC>	LATIN SMALL LETTER NJ
4259	<A<>	<U01CD>	LATIN CAPITAL LETTER A WITH CARON
4260	<a<>	<U01CE>	LATIN SMALL LETTER A WITH CARON
4261	<I<>	<U01CF>	LATIN CAPITAL LETTER I WITH CARON
4262	<i<>	<U01D0>	LATIN SMALL LETTER I WITH CARON
4263	<O<>	<U01D1>	LATIN CAPITAL LETTER O WITH CARON
4264	<o<>	<U01D2>	LATIN SMALL LETTER O WITH CARON
4265	<U<>	<U01D3>	LATIN CAPITAL LETTER U WITH CARON
4266	<u<>	<U01D4>	LATIN SMALL LETTER U WITH CARON
4267	<U: ->	<U01D5>	LATIN CAPITAL LETTER U WITH DIAERESIS AND MACRON
4268	<u: ->	<U01D6>	LATIN SMALL LETTER U WITH DIAERESIS AND MACRON
4269	<U: ' >	<U01D7>	LATIN CAPITAL LETTER U WITH DIAERESIS AND ACUTE
4270	<u: ' >	<U01D8>	LATIN SMALL LETTER U WITH DIAERESIS AND ACUTE
4271	<U: <>	<U01D9>	LATIN CAPITAL LETTER U WITH DIAERESIS AND CARON
4272	<u: <>	<U01DA>	LATIN SMALL LETTER U WITH DIAERESIS AND CARON
4273	<U: ! >	<U01DB>	LATIN CAPITAL LETTER U WITH DIAERESIS AND GRAVE
4274	<u: ! >	<U01DC>	LATIN SMALL LETTER U WITH DIAERESIS AND GRAVE
4275	<e1>	<U01DD>	LATIN SMALL LETTER TURNED E
4276	<A1>	<U01DE>	LATIN CAPITAL LETTER A WITH DIAERESIS AND MACRON
4277	<a1>	<U01DF>	LATIN SMALL LETTER A WITH DIAERESIS AND MACRON
4278	<A7>	<U01E0>	LATIN CAPITAL LETTER A WITH DOT ABOVE AND MACRON
4279	<a7>	<U01E1>	LATIN SMALL LETTER A WITH DOT ABOVE AND MACRON
4280	<A3>	<U01E2>	LATIN CAPITAL LETTER AE WITH MACRON (ash)
4281	<a3>	<U01E3>	LATIN SMALL LETTER AE WITH MACRON (ash)
4282	<G//>	<U01E4>	LATIN CAPITAL LETTER G WITH STROKE
4283	<g//>	<U01E5>	LATIN SMALL LETTER G WITH STROKE
4284	<G<>	<U01E6>	LATIN CAPITAL LETTER G WITH CARON
4285	<g<>	<U01E7>	LATIN SMALL LETTER G WITH CARON
4286	<K<>	<U01E8>	LATIN CAPITAL LETTER K WITH CARON
4287	<k<>	<U01E9>	LATIN SMALL LETTER K WITH CARON
4288	<O; >	<U01EA>	LATIN CAPITAL LETTER O WITH OGONEK
4289	<o; >	<U01EB>	LATIN SMALL LETTER O WITH OGONEK
4290	<O1>	<U01EC>	LATIN CAPITAL LETTER O WITH OGONEK AND MACRON
4291	<o1>	<U01ED>	LATIN SMALL LETTER O WITH OGONEK AND MACRON
4292	<EZ>	<U01EE>	LATIN CAPITAL LETTER EZH WITH CARON
4293	<ez>	<U01EF>	LATIN SMALL LETTER EZH WITH CARON
4294	<j<>	<U01F0>	LATIN SMALL LETTER J WITH CARON
4295	<DZ3>	<U01F1>	LATIN CAPITAL LETTER DZ
4296	<Dz3>	<U01F2>	LATIN CAPITAL LETTER D WITH SMALL LETTER Z
4297	<dz3>	<U01F3>	LATIN SMALL LETTER DZ
4298	<G' >	<U01F4>	LATIN CAPITAL LETTER G WITH ACUTE
4299	<g' >	<U01F5>	LATIN SMALL LETTER G WITH ACUTE
4300	<AA' >	<U01FA>	LATIN CAPITAL LETTER A WITH RING ABOVE AND ACUTE
4301	<aa' >	<U01FB>	LATIN SMALL LETTER A WITH RING ABOVE AND ACUTE
4302	<AE' >	<U01FC>	LATIN CAPITAL LETTER AE WITH ACUTE (ash)
4303	<ae' >	<U01FD>	LATIN SMALL LETTER AE WITH ACUTE (ash)
4304	<O//' >	<U01FE>	LATIN CAPITAL LETTER O WITH STROKE AND ACUTE
4305	<o//' >	<U01FF>	LATIN SMALL LETTER O WITH STROKE AND ACUTE
4306	<A!! >	<U0200>	LATIN CAPITAL LETTER A WITH DOUBLE GRAVE
4307	<a!! >	<U0201>	LATIN SMALL LETTER A WITH DOUBLE GRAVE
4308	<A) >	<U0202>	LATIN CAPITAL LETTER A WITH INVERTED BREVE
4309	<a) >	<U0203>	LATIN SMALL LETTER A WITH INVERTED BREVE
4310	<E!! >	<U0204>	LATIN CAPITAL LETTER E WITH DOUBLE GRAVE
4311	<e!! >	<U0205>	LATIN SMALL LETTER E WITH DOUBLE GRAVE
4312	<E) >	<U0206>	LATIN CAPITAL LETTER E WITH INVERTED BREVE
4313	<e) >	<U0207>	LATIN SMALL LETTER E WITH INVERTED BREVE
4314	<I!! >	<U0208>	LATIN CAPITAL LETTER I WITH DOUBLE GRAVE
4315	<i!! >	<U0209>	LATIN SMALL LETTER I WITH DOUBLE GRAVE
4316	<I) >	<U020A>	LATIN CAPITAL LETTER I WITH INVERTED BREVE
4317	<i) >	<U020B>	LATIN SMALL LETTER I WITH INVERTED BREVE
4318	<O!! >	<U020C>	LATIN CAPITAL LETTER O WITH DOUBLE GRAVE

4319	<o!!>	<U020D>	LATIN SMALL LETTER O WITH DOUBLE GRAVE
4320	<O>	<U020E>	LATIN CAPITAL LETTER O WITH INVERTED BREVE
4321	<o>	<U020F>	LATIN SMALL LETTER O WITH INVERTED BREVE
4322	<R!!>	<U0210>	LATIN CAPITAL LETTER R WITH DOUBLE GRAVE
4323	<r!!>	<U0211>	LATIN SMALL LETTER R WITH DOUBLE GRAVE
4324	<R>	<U0212>	LATIN CAPITAL LETTER R WITH INVERTED BREVE
4325	<r>	<U0213>	LATIN SMALL LETTER R WITH INVERTED BREVE
4326	<U!!>	<U0214>	LATIN CAPITAL LETTER U WITH DOUBLE GRAVE
4327	<u!!>	<U0215>	LATIN SMALL LETTER U WITH DOUBLE GRAVE
4328	<U>	<U0216>	LATIN CAPITAL LETTER U WITH INVERTED BREVE
4329	<u>	<U0217>	LATIN SMALL LETTER U WITH INVERTED BREVE
4330	<r1>	<U027C>	LATIN SMALL LETTER R WITH LONG LEG
4331	<ed>	<U0292>	LATIN SMALL LETTER EZH
4332	<¡S>	<U02BB>	MODIFIER LETTER TURNED COMMA
4333	<l/>	<U02C6>	MODIFIER LETTER CIRCUMFLEX ACCENT
4334	<'<>	<U02C7>	CARON (Mandarin Chinese third tone)
4335	<l->	<U02C9>	MODIFIER LETTER MACRON (Mandarin Chinese first tone)
4336	<l!>	<U02CB>	MODIFIER LETTER GRAVE ACCENT (Mandarin Chinese fourth tone)
4337	<'(>	<U02D8>	BREVE
4338	<'.>	<U02D9>	DOT ABOVE (Mandarin Chinese light tone)
4339	<'0>	<U02DA>	RING ABOVE
4340	<'>	<U02DB>	OGONEK
4341	<l?>	<U02DC>	SMALL TILDE
4342	<' " >	<U02DD>	DOUBLE ACUTE ACCENT
4343	<'G>	<U0374>	GREEK NUMERAL SIGN (Dexia keraia)
4344	<,G>	<U0375>	GREEK LOWER NUMERAL SIGN (Aristeri keraia)
4345	<j3>	<U037A>	GREEK YPOGEGRAMMENI
4346	<?%>	<U037E>	GREEK QUESTION MARK (Erotimatiko)
4347	<' * >	<U0384>	GREEK TONOS
4348	<' % >	<U0385>	GREEK DIALYTIKA TONOS
4349	<A%>	<U0386>	GREEK CAPITAL LETTER ALPHA WITH TONOS
4350	<. * >	<U0387>	GREEK ANO TELEIA
4351	<E%>	<U0388>	GREEK CAPITAL LETTER EPSILON WITH TONOS
4352	<Y%>	<U0389>	GREEK CAPITAL LETTER ETA WITH TONOS
4353	<I%>	<U038A>	GREEK CAPITAL LETTER IOTA WITH TONOS
4354	<O%>	<U038C>	GREEK CAPITAL LETTER OMICRON WITH TONOS
4355	<U%>	<U038E>	GREEK CAPITAL LETTER UPSILON WITH TONOS
4356	<W%>	<U038F>	GREEK CAPITAL LETTER OMEGA WITH TONOS
4357	<i3>	<U0390>	GREEK SMALL LETTER IOTA WITH DIALYTIKA AND TONOS
4358	<A* >	<U0391>	GREEK CAPITAL LETTER ALPHA
4359	<B* >	<U0392>	GREEK CAPITAL LETTER BETA
4360	<G* >	<U0393>	GREEK CAPITAL LETTER GAMMA
4361	<D* >	<U0394>	GREEK CAPITAL LETTER DELTA
4362	<E* >	<U0395>	GREEK CAPITAL LETTER EPSILON
4363	<Z* >	<U0396>	GREEK CAPITAL LETTER ZETA
4364	<Y* >	<U0397>	GREEK CAPITAL LETTER ETA
4365	<H* >	<U0398>	GREEK CAPITAL LETTER THETA
4366	<I* >	<U0399>	GREEK CAPITAL LETTER IOTA
4367	<K* >	<U039A>	GREEK CAPITAL LETTER KAPPA
4368	<L* >	<U039B>	GREEK CAPITAL LETTER LAMDA
4369	<M* >	<U039C>	GREEK CAPITAL LETTER MU
4370	<N* >	<U039D>	GREEK CAPITAL LETTER NU
4371	<C* >	<U039E>	GREEK CAPITAL LETTER XI
4372	<O* >	<U039F>	GREEK CAPITAL LETTER OMICRON
4373	<P* >	<U03A0>	GREEK CAPITAL LETTER PI
4374	<R* >	<U03A1>	GREEK CAPITAL LETTER RHO
4375	<S* >	<U03A3>	GREEK CAPITAL LETTER SIGMA
4376	<T* >	<U03A4>	GREEK CAPITAL LETTER TAU
4377	<U* >	<U03A5>	GREEK CAPITAL LETTER UPSILON
4378	<F* >	<U03A6>	GREEK CAPITAL LETTER PHI
4379	<X* >	<U03A7>	GREEK CAPITAL LETTER CHI
4380	<Q* >	<U03A8>	GREEK CAPITAL LETTER PSI
4381	<W* >	<U03A9>	GREEK CAPITAL LETTER OMEGA
4382	<J* >	<U03AA>	GREEK CAPITAL LETTER IOTA WITH DIALYTIKA
4383	<V* >	<U03AB>	GREEK CAPITAL LETTER UPSILON WITH DIALYTIKA
4384	<a%>	<U03AC>	GREEK SMALL LETTER ALPHA WITH TONOS
4385	<e%>	<U03AD>	GREEK SMALL LETTER EPSILON WITH TONOS
4386	<y%>	<U03AE>	GREEK SMALL LETTER ETA WITH TONOS
4387	<i%>	<U03AF>	GREEK SMALL LETTER IOTA WITH TONOS
4388	<u3>	<U03B0>	GREEK SMALL LETTER UPSILON WITH DIALYTIKA AND TONOS
4389	<a* >	<U03B1>	GREEK SMALL LETTER ALPHA
4390	<b* >	<U03B2>	GREEK SMALL LETTER BETA
4391	<g* >	<U03B3>	GREEK SMALL LETTER GAMMA
4392	<d* >	<U03B4>	GREEK SMALL LETTER DELTA
4393	<e* >	<U03B5>	GREEK SMALL LETTER EPSILON
4394	<z* >	<U03B6>	GREEK SMALL LETTER ZETA
4395	<y* >	<U03B7>	GREEK SMALL LETTER ETA
4396	<h* >	<U03B8>	GREEK SMALL LETTER THETA
4397	<i* >	<U03B9>	GREEK SMALL LETTER IOTA
4398	<k* >	<U03BA>	GREEK SMALL LETTER KAPPA
4399	<l* >	<U03BB>	GREEK SMALL LETTER LAMDA
4400	<m* >	<U03BC>	GREEK SMALL LETTER MU
4401	<n* >	<U03BD>	GREEK SMALL LETTER NU
4402	<c* >	<U03BE>	GREEK SMALL LETTER XI
4403	<o* >	<U03BF>	GREEK SMALL LETTER OMICRON
4404	<p* >	<U03C0>	GREEK SMALL LETTER PI
4405	<r* >	<U03C1>	GREEK SMALL LETTER RHO
4406	<*s>	<U03C2>	GREEK SMALL LETTER FINAL SIGMA
4407	<s* >	<U03C3>	GREEK SMALL LETTER SIGMA

4408	<t*>	<U03C4>	GREEK SMALL LETTER TAU
4409	<u*>	<U03C5>	GREEK SMALL LETTER UPSILON
4410	<f*>	<U03C6>	GREEK SMALL LETTER PHI
4411	<x*>	<U03C7>	GREEK SMALL LETTER CHI
4412	<q*>	<U03C8>	GREEK SMALL LETTER PSI
4413	<w*>	<U03C9>	GREEK SMALL LETTER OMEGA
4414	<j*>	<U03CA>	GREEK SMALL LETTER IOTA WITH DIALYTIKA
4415	<v*>	<U03CB>	GREEK SMALL LETTER UPSILON WITH DIALYTIKA
4416	<o%>	<U03CC>	GREEK SMALL LETTER OMICRON WITH TONOS
4417	<u%>	<U03CD>	GREEK SMALL LETTER UPSILON WITH TONOS
4418	<w%>	<U03CE>	GREEK SMALL LETTER OMEGA WITH TONOS
4419	<b3>	<U03D0>	GREEK BETA SYMBOL
4420	<T3>	<U03DA>	GREEK LETTER STIGMA
4421	<M3>	<U03DC>	GREEK LETTER DIGAMMA
4422	<K3>	<U03DE>	GREEK LETTER KOPPA
4423	<P3>	<U03E0>	GREEK LETTER SAMPI
4424	<IO>	<U0401>	CYRILLIC CAPITAL LETTER IO
4425	<D%>	<U0402>	CYRILLIC CAPITAL LETTER DJE (Serbocroatian)
4426	<G%>	<U0403>	CYRILLIC CAPITAL LETTER GJE
4427	<IE>	<U0404>	CYRILLIC CAPITAL LETTER UKRAINIAN IE
4428	<DS>	<U0405>	CYRILLIC CAPITAL LETTER DZE
4429	<II>	<U0406>	CYRILLIC CAPITAL LETTER BYELORUSSIAN-UKRAINIAN I
4430	<YI>	<U0407>	CYRILLIC CAPITAL LETTER YI (Ukrainian)
4431	<J%>	<U0408>	CYRILLIC CAPITAL LETTER JE
4432	<LJ>	<U0409>	CYRILLIC CAPITAL LETTER LJJE
4433	<NJ>	<U040A>	CYRILLIC CAPITAL LETTER NJJE
4434	<Ts>	<U040B>	CYRILLIC CAPITAL LETTER TSHE (Serbocroatian)
4435	<KJ>	<U040C>	CYRILLIC CAPITAL LETTER KJE
4436	<V%>	<U040E>	CYRILLIC CAPITAL LETTER SHORT U (Byelorussian)
4437	<DZ>	<U040F>	CYRILLIC CAPITAL LETTER DZHE
4438	<A>	<U0410>	CYRILLIC CAPITAL LETTER A
4439	<B>	<U0411>	CYRILLIC CAPITAL LETTER BE
4440	<V=>	<U0412>	CYRILLIC CAPITAL LETTER VE
4441	<G=>	<U0413>	CYRILLIC CAPITAL LETTER GHE
4442	<D=>	<U0414>	CYRILLIC CAPITAL LETTER DE
4443	<E=>	<U0415>	CYRILLIC CAPITAL LETTER IE
4444	<Z%>	<U0416>	CYRILLIC CAPITAL LETTER ZHE
4445	<Z=>	<U0417>	CYRILLIC CAPITAL LETTER ZE
4446	<I=>	<U0418>	CYRILLIC CAPITAL LETTER I
4447	<J=>	<U0419>	CYRILLIC CAPITAL LETTER SHORT I
4448	<K=>	<U041A>	CYRILLIC CAPITAL LETTER KA
4449	<L=>	<U041B>	CYRILLIC CAPITAL LETTER EL
4450	<M=>	<U041C>	CYRILLIC CAPITAL LETTER EM
4451	<N=>	<U041D>	CYRILLIC CAPITAL LETTER EN
4452	<O=>	<U041E>	CYRILLIC CAPITAL LETTER O
4453	<P=>	<U041F>	CYRILLIC CAPITAL LETTER PE
4454	<R=>	<U0420>	CYRILLIC CAPITAL LETTER ER
4455	<S=>	<U0421>	CYRILLIC CAPITAL LETTER ES
4456	<T=>	<U0422>	CYRILLIC CAPITAL LETTER TE
4457	<U=>	<U0423>	CYRILLIC CAPITAL LETTER U
4458	<F=>	<U0424>	CYRILLIC CAPITAL LETTER EF
4459	<H=>	<U0425>	CYRILLIC CAPITAL LETTER HA
4460	<C=>	<U0426>	CYRILLIC CAPITAL LETTER TSE
4461	<C%>	<U0427>	CYRILLIC CAPITAL LETTER CHE
4462	<S%>	<U0428>	CYRILLIC CAPITAL LETTER SHA
4463	<Sc>	<U0429>	CYRILLIC CAPITAL LETTER SHCHA
4464	<=>	<U042A>	CYRILLIC CAPITAL LETTER HARD SIGN
4465	<Y=>	<U042B>	CYRILLIC CAPITAL LETTER YERU
4466	<%>	<U042C>	CYRILLIC CAPITAL LETTER SOFT SIGN
4467	<JE>	<U042D>	CYRILLIC CAPITAL LETTER E
4468	<JU>	<U042E>	CYRILLIC CAPITAL LETTER YU
4469	<JA>	<U042F>	CYRILLIC CAPITAL LETTER YA
4470	<a=>	<U0430>	CYRILLIC SMALL LETTER A
4471	<b=>	<U0431>	CYRILLIC SMALL LETTER BE
4472	<v=>	<U0432>	CYRILLIC SMALL LETTER VE
4473	<g=>	<U0433>	CYRILLIC SMALL LETTER GHE
4474	<d=>	<U0434>	CYRILLIC SMALL LETTER DE
4475	<e=>	<U0435>	CYRILLIC SMALL LETTER IE
4476	<z%>	<U0436>	CYRILLIC SMALL LETTER ZHE
4477	<z=>	<U0437>	CYRILLIC SMALL LETTER ZE
4478	<i=>	<U0438>	CYRILLIC SMALL LETTER I
4479	<j=>	<U0439>	CYRILLIC SMALL LETTER SHORT I
4480	<k=>	<U043A>	CYRILLIC SMALL LETTER KA
4481	<l=>	<U043B>	CYRILLIC SMALL LETTER EL
4482	<m=>	<U043C>	CYRILLIC SMALL LETTER EM
4483	<n=>	<U043D>	CYRILLIC SMALL LETTER EN
4484	<o=>	<U043E>	CYRILLIC SMALL LETTER O
4485	<p=>	<U043F>	CYRILLIC SMALL LETTER PE
4486	<r=>	<U0440>	CYRILLIC SMALL LETTER ER
4487	<s=>	<U0441>	CYRILLIC SMALL LETTER ES
4488	<t=>	<U0442>	CYRILLIC SMALL LETTER TE
4489	<u=>	<U0443>	CYRILLIC SMALL LETTER U
4490	<f=>	<U0444>	CYRILLIC SMALL LETTER EF
4491	<h=>	<U0445>	CYRILLIC SMALL LETTER HA
4492	<c=>	<U0446>	CYRILLIC SMALL LETTER TSE
4493	<c%>	<U0447>	CYRILLIC SMALL LETTER CHE
4494	<s%>	<U0448>	CYRILLIC SMALL LETTER SHA

4495	<sc>	<U0449>	CYRILLIC SMALL LETTER SHCHA
4496	<=/>	<U044A>	CYRILLIC SMALL LETTER HARD SIGN
4497	<y=>	<U044B>	CYRILLIC SMALL LETTER YERU
4498	<%'/>	<U044C>	CYRILLIC SMALL LETTER SOFT SIGN
4499	<je>	<U044D>	CYRILLIC SMALL LETTER E
4500	<ju>	<U044E>	CYRILLIC SMALL LETTER YU
4501	<ja>	<U044F>	CYRILLIC SMALL LETTER YA
4502	<io>	<U0451>	CYRILLIC SMALL LETTER IO
4503	<d%>	<U0452>	CYRILLIC SMALL LETTER DJE (Serbocroatian)
4504	<g%>	<U0453>	CYRILLIC SMALL LETTER GJE
4505	<ie>	<U0454>	CYRILLIC SMALL LETTER UKRAINIAN IE
4506	<ds>	<U0455>	CYRILLIC SMALL LETTER DZE
4507	<ii>	<U0456>	CYRILLIC SMALL LETTER BYELORUSSIAN-UKRAINIAN I
4508	<yi>	<U0457>	CYRILLIC SMALL LETTER YI (Ukrainian)
4509	<j%>	<U0458>	CYRILLIC SMALL LETTER JE
4510	<lj>	<U0459>	CYRILLIC SMALL LETTER LJE
4511	<nj>	<U045A>	CYRILLIC SMALL LETTER NJE
4512	<ts>	<U045B>	CYRILLIC SMALL LETTER TSHE (Serbocroatian)
4513	<kj>	<U045C>	CYRILLIC SMALL LETTER KJE
4514	<v%>	<U045E>	CYRILLIC SMALL LETTER SHORT U (Byelorussian)
4515	<dz>	<U045F>	CYRILLIC SMALL LETTER DZHE
4516	<Y3>	<U0462>	CYRILLIC CAPITAL LETTER YAT
4517	<y3>	<U0463>	CYRILLIC SMALL LETTER YAT
4518	<O3>	<U046A>	CYRILLIC CAPITAL LETTER BIG YUS
4519	<o3>	<U046B>	CYRILLIC SMALL LETTER BIG YUS
4520	<F3>	<U0472>	CYRILLIC CAPITAL LETTER FITA
4521	<f3>	<U0473>	CYRILLIC SMALL LETTER FITA
4522	<V3>	<U0474>	CYRILLIC CAPITAL LETTER IZHITSA
4523	<v3>	<U0475>	CYRILLIC SMALL LETTER IZHITSA
4524	<C3>	<U0480>	CYRILLIC CAPITAL LETTER KOPPA
4525	<c3>	<U0481>	CYRILLIC SMALL LETTER KOPPA
4526	<G3>	<U0490>	CYRILLIC CAPITAL LETTER GHE WITH UPTURN
4527	<g3>	<U0491>	CYRILLIC SMALL LETTER GHE WITH UPTURN
4528	<A+>	<U05D0>	HEBREW LETTER ALEF
4529	<B+>	<U05D1>	HEBREW LETTER BET
4530	<G+>	<U05D2>	HEBREW LETTER GIMEL
4531	<D+>	<U05D3>	HEBREW LETTER DALET
4532	<H+>	<U05D4>	HEBREW LETTER HE
4533	<W+>	<U05D5>	HEBREW LETTER VAV
4534	<Z+>	<U05D6>	HEBREW LETTER ZAYIN
4535	<X+>	<U05D7>	HEBREW LETTER HET
4536	<Tj>	<U05D8>	HEBREW LETTER TET
4537	<J+>	<U05D9>	HEBREW LETTER YOD
4538	<K%>	<U05DA>	HEBREW LETTER FINAL KAF
4539	<K+>	<U05DB>	HEBREW LETTER KAF
4540	<L+>	<U05DC>	HEBREW LETTER LAMED
4541	<M%>	<U05DD>	HEBREW LETTER FINAL MEM
4542	<M+>	<U05DE>	HEBREW LETTER MEM
4543	<N%>	<U05DF>	HEBREW LETTER FINAL NUN
4544	<N+>	<U05E0>	HEBREW LETTER NUN
4545	<S+>	<U05E1>	HEBREW LETTER SAMEKH
4546	<E+>	<U05E2>	HEBREW LETTER AYIN
4547	<P%>	<U05E3>	HEBREW LETTER FINAL PE
4548	<P+>	<U05E4>	HEBREW LETTER PE
4549	<Zj>	<U05E5>	HEBREW LETTER FINAL TSADI
4550	<ZJ>	<U05E6>	HEBREW LETTER TSADI
4551	<Q+>	<U05E7>	HEBREW LETTER QOF
4552	<R+>	<U05E8>	HEBREW LETTER RESH
4553	<Sh>	<U05E9>	HEBREW LETTER SHIN
4554	<T+>	<U05EA>	HEBREW LETTER TAV
4555	<,+>	<U060C>	ARABIC COMMA
4556	< ; +>	<U061B>	ARABIC SEMICOLON
4557	<?+>	<U061F>	ARABIC QUESTION MARK
4558	<H'>	<U0621>	ARABIC LETTER HAMZA
4559	<aM>	<U0622>	ARABIC LETTER ALEF WITH MADDA ABOVE
4560	<aH>	<U0623>	ARABIC LETTER ALEF WITH HAMZA ABOVE
4561	<wH>	<U0624>	ARABIC LETTER WAW WITH HAMZA ABOVE
4562	<ah>	<U0625>	ARABIC LETTER ALEF WITH HAMZA BELOW
4563	<yH>	<U0626>	ARABIC LETTER YEH WITH HAMZA ABOVE
4564	<a+>	<U0627>	ARABIC LETTER ALEF
4565	<b+>	<U0628>	ARABIC LETTER BEH
4566	<tm>	<U0629>	ARABIC LETTER TEH MARBUTA
4567	<t+>	<U062A>	ARABIC LETTER TEH
4568	<tk>	<U062B>	ARABIC LETTER THEH
4569	<g+>	<U062C>	ARABIC LETTER JEEM
4570	<hk>	<U062D>	ARABIC LETTER HAH
4571	<x+>	<U062E>	ARABIC LETTER KHAH
4572	<d+>	<U062F>	ARABIC LETTER DAL
4573	<dk>	<U0630>	ARABIC LETTER THAL
4574	<r+>	<U0631>	ARABIC LETTER REH
4575	<z+>	<U0632>	ARABIC LETTER ZAIN
4576	<s+>	<U0633>	ARABIC LETTER SEEN
4577	<sn>	<U0634>	ARABIC LETTER SHEEN
4578	<c+>	<U0635>	ARABIC LETTER SAD
4579	<dd>	<U0636>	ARABIC LETTER DAD
4580	<tj>	<U0637>	ARABIC LETTER TAH
4581	<zH>	<U0638>	ARABIC LETTER ZAH
4582	<e+>	<U0639>	ARABIC LETTER AIN
4583	<i+>	<U063A>	ARABIC LETTER GHAIN

4584	<++>	<U0640>	ARABIC TATWEEL
4585	<f+>	<U0641>	ARABIC LETTER FEH
4586	<q+>	<U0642>	ARABIC LETTER QAF
4587	<k+>	<U0643>	ARABIC LETTER KAF
4588	<l+>	<U0644>	ARABIC LETTER LAM
4589	<m+>	<U0645>	ARABIC LETTER MEEM
4590	<n+>	<U0646>	ARABIC LETTER NOON
4591	<h+>	<U0647>	ARABIC LETTER HEH
4592	<w+>	<U0648>	ARABIC LETTER WAW
4593	<j+>	<U0649>	ARABIC LETTER ALEF MAKSURA
4594	<y+>	<U064A>	ARABIC LETTER YEH
4595	<:+>	<U064B>	ARABIC FATHATAN
4596	<"+>	<U064C>	ARABIC DAMMATAN
4597	<==>	<U064D>	ARABIC KASRATAN
4598	<//+>	<U064E>	ARABIC FATHA
4599	<' +>	<U064F>	ARABIC DAMMA
4600	<1+>	<U0650>	ARABIC KASRA
4601	<3+>	<U0651>	ARABIC SHADDA
4602	<0+>	<U0652>	ARABIC SUKUN
4603	<0a>	<U0660>	ARABIC-INDIC DIGIT ZERO
4604	<1a>	<U0661>	ARABIC-INDIC DIGIT ONE
4605	<2a>	<U0662>	ARABIC-INDIC DIGIT TWO
4606	<3a>	<U0663>	ARABIC-INDIC DIGIT THREE
4607	<4a>	<U0664>	ARABIC-INDIC DIGIT FOUR
4608	<5a>	<U0665>	ARABIC-INDIC DIGIT FIVE
4609	<6a>	<U0666>	ARABIC-INDIC DIGIT SIX
4610	<7a>	<U0667>	ARABIC-INDIC DIGIT SEVEN
4611	<8a>	<U0668>	ARABIC-INDIC DIGIT EIGHT
4612	<9a>	<U0669>	ARABIC-INDIC DIGIT NINE
4613	<aS>	<U0670>	ARABIC LETTER SUPERSCRIPIT ALEF
4614	<p+>	<U067E>	ARABIC LETTER PEH
4615	<hH>	<U0681>	ARABIC LETTER HAH WITH HAMZA ABOVE
4616	<tc>	<U0686>	ARABIC LETTER TCHEH
4617	<zj>	<U0698>	ARABIC LETTER JEH
4618	<v+>	<U06A4>	ARABIC LETTER VEH
4619	<gf>	<U06AF>	ARABIC LETTER GAF
4620	<A-0>	<U1E00>	LATIN CAPITAL LETTER A WITH RING BELOW
4621	<a-0>	<U1E01>	LATIN SMALL LETTER A WITH RING BELOW
4622	<B.>	<U1E02>	LATIN CAPITAL LETTER B WITH DOT ABOVE
4623	<b.>	<U1E03>	LATIN SMALL LETTER B WITH DOT ABOVE
4624	<B-.>	<U1E04>	LATIN CAPITAL LETTER B WITH DOT BELOW
4625	<b-.>	<U1E05>	LATIN SMALL LETTER B WITH DOT BELOW
4626	<B_>	<U1E06>	LATIN CAPITAL LETTER B WITH LINE BELOW
4627	<b_>	<U1E07>	LATIN SMALL LETTER B WITH LINE BELOW
4628	<C,'>	<U1E08>	LATIN CAPITAL LETTER C WITH CEDILLA AND ACUTE
4629	<c,'>	<U1E09>	LATIN SMALL LETTER C WITH CEDILLA AND ACUTE
4630	<D.>	<U1E0A>	LATIN CAPITAL LETTER D WITH DOT ABOVE
4631	<d.>	<U1E0B>	LATIN SMALL LETTER D WITH DOT ABOVE
4632	<D-.>	<U1E0C>	LATIN CAPITAL LETTER D WITH DOT BELOW
4633	<d-.>	<U1E0D>	LATIN SMALL LETTER D WITH DOT BELOW
4634	<D_>	<U1E0E>	LATIN CAPITAL LETTER D WITH LINE BELOW
4635	<d_>	<U1E0F>	LATIN SMALL LETTER D WITH LINE BELOW
4636	<D,>	<U1E10>	LATIN CAPITAL LETTER D WITH CEDILLA
4637	<d,>	<U1E11>	LATIN SMALL LETTER D WITH CEDILLA
4638	<D-/>>	<U1E12>	LATIN CAPITAL LETTER D WITH CIRCUMFLEX BELOW
4639	<d-/>>	<U1E13>	LATIN SMALL LETTER D WITH CIRCUMFLEX BELOW
4640	<E-!>	<U1E14>	LATIN CAPITAL LETTER E WITH MACRON AND GRAVE
4641	<e-!>	<U1E15>	LATIN SMALL LETTER E WITH MACRON AND GRAVE
4642	<E-'>	<U1E16>	LATIN CAPITAL LETTER E WITH MACRON AND ACUTE
4643	<e-'>	<U1E17>	LATIN SMALL LETTER E WITH MACRON AND ACUTE
4644	<E-/>>	<U1E18>	LATIN CAPITAL LETTER E WITH CIRCUMFLEX BELOW
4645	<e-/>>	<U1E19>	LATIN SMALL LETTER E WITH CIRCUMFLEX BELOW
4646	<E-?>	<U1E1A>	LATIN CAPITAL LETTER E WITH TILDE BELOW
4647	<e-?>	<U1E1B>	LATIN SMALL LETTER E WITH TILDE BELOW
4648	<E,( >	<U1E1C>	LATIN CAPITAL LETTER E WITH CEDILLA AND BREVE
4649	<e,( >	<U1E1D>	LATIN SMALL LETTER E WITH CEDILLA AND BREVE
4650	<F.>	<U1E1E>	LATIN CAPITAL LETTER F WITH DOT ABOVE
4651	<f.>	<U1E1F>	LATIN SMALL LETTER F WITH DOT ABOVE
4652	<G->	<U1E20>	LATIN CAPITAL LETTER G WITH MACRON
4653	<g->	<U1E21>	LATIN SMALL LETTER G WITH MACRON
4654	<H.>	<U1E22>	LATIN CAPITAL LETTER H WITH DOT ABOVE
4655	<h.>	<U1E23>	LATIN SMALL LETTER H WITH DOT ABOVE
4656	<H-.>	<U1E24>	LATIN CAPITAL LETTER H WITH DOT BELOW
4657	<h-.>	<U1E25>	LATIN SMALL LETTER H WITH DOT BELOW
4658	<H:>	<U1E26>	LATIN CAPITAL LETTER H WITH DIAERESIS
4659	<h:>	<U1E27>	LATIN SMALL LETTER H WITH DIAERESIS
4660	<H,>	<U1E28>	LATIN CAPITAL LETTER H WITH CEDILLA
4661	<h,>	<U1E29>	LATIN SMALL LETTER H WITH CEDILLA
4662	<H-( >	<U1E2A>	LATIN CAPITAL LETTER H WITH BREVE BELOW
4663	<h-( >	<U1E2B>	LATIN SMALL LETTER H WITH BREVE BELOW
4664	<I-?>	<U1E2C>	LATIN CAPITAL LETTER I WITH TILDE BELOW
4665	<i-?>	<U1E2D>	LATIN SMALL LETTER I WITH TILDE BELOW
4666	<I:'>	<U1E2E>	LATIN CAPITAL LETTER I WITH DIAERESIS AND ACUTE
4667	<i:'>	<U1E2F>	LATIN SMALL LETTER I WITH DIAERESIS AND ACUTE
4668	<K'>	<U1E30>	LATIN CAPITAL LETTER K WITH ACUTE
4669	<k'>	<U1E31>	LATIN SMALL LETTER K WITH ACUTE
4670	<K-.>	<U1E32>	LATIN CAPITAL LETTER K WITH DOT BELOW

4671	<k-.>	<U1E33>	LATIN SMALL LETTER K WITH DOT BELOW
4672	<K_>	<U1E34>	LATIN CAPITAL LETTER K WITH LINE BELOW
4673	<k_>	<U1E35>	LATIN SMALL LETTER K WITH LINE BELOW
4674	<L-.>	<U1E36>	LATIN CAPITAL LETTER L WITH DOT BELOW
4675	<l-.>	<U1E37>	LATIN SMALL LETTER L WITH DOT BELOW
4676	<L--.>	<U1E38>	LATIN CAPITAL LETTER L WITH DOT BELOW AND MACRON
4677	<l--.>	<U1E39>	LATIN SMALL LETTER L WITH DOT BELOW AND MACRON
4678	<L_>	<U1E3A>	LATIN CAPITAL LETTER L WITH LINE BELOW
4679	<l_>	<U1E3B>	LATIN SMALL LETTER L WITH LINE BELOW
4680	<L-/>	<U1E3C>	LATIN CAPITAL LETTER L WITH CIRCUMFLEX BELOW
4681	<l-/>	<U1E3D>	LATIN SMALL LETTER L WITH CIRCUMFLEX BELOW
4682	<M'>	<U1E3E>	LATIN CAPITAL LETTER M WITH ACUTE
4683	<m'>	<U1E3F>	LATIN SMALL LETTER M WITH ACUTE
4684	<M.>	<U1E40>	LATIN CAPITAL LETTER M WITH DOT ABOVE
4685	<m.>	<U1E41>	LATIN SMALL LETTER M WITH DOT ABOVE
4686	<M-.>	<U1E42>	LATIN CAPITAL LETTER M WITH DOT BELOW
4687	<m-.>	<U1E43>	LATIN SMALL LETTER M WITH DOT BELOW
4688	<N.>	<U1E44>	LATIN CAPITAL LETTER N WITH DOT ABOVE
4689	<n.>	<U1E45>	LATIN SMALL LETTER N WITH DOT ABOVE
4690	<N-.>	<U1E46>	LATIN CAPITAL LETTER N WITH DOT BELOW
4691	<n-.>	<U1E47>	LATIN SMALL LETTER N WITH DOT BELOW
4692	<N_>	<U1E48>	LATIN CAPITAL LETTER N WITH LINE BELOW
4693	<n_>	<U1E49>	LATIN SMALL LETTER N WITH LINE BELOW
4694	<N-/>	<U1E4A>	LATIN CAPITAL LETTER N WITH CIRCUMFLEX BELOW
4695	<n-/>	<U1E4B>	LATIN SMALL LETTER N WITH CIRCUMFLEX BELOW
4696	<O?'>	<U1E4C>	LATIN CAPITAL LETTER O WITH TILDE AND ACUTE
4697	<o?'>	<U1E4D>	LATIN SMALL LETTER O WITH TILDE AND ACUTE
4698	<O?:'>	<U1E4E>	LATIN CAPITAL LETTER O WITH TILDE AND DIAERESIS
4699	<o?:'>	<U1E4F>	LATIN SMALL LETTER O WITH TILDE AND DIAERESIS
4700	<O-!>	<U1E50>	LATIN CAPITAL LETTER O WITH MACRON AND GRAVE
4701	<o-!>	<U1E51>	LATIN SMALL LETTER O WITH MACRON AND GRAVE
4702	<O-'>	<U1E52>	LATIN CAPITAL LETTER O WITH MACRON AND ACUTE
4703	<o-'>	<U1E53>	LATIN SMALL LETTER O WITH MACRON AND ACUTE
4704	<P'>	<U1E54>	LATIN CAPITAL LETTER P WITH ACUTE
4705	<p'>	<U1E55>	LATIN SMALL LETTER P WITH ACUTE
4706	<P.>	<U1E56>	LATIN CAPITAL LETTER P WITH DOT ABOVE
4707	<p.>	<U1E57>	LATIN SMALL LETTER P WITH DOT ABOVE
4708	<R.>	<U1E58>	LATIN CAPITAL LETTER R WITH DOT ABOVE
4709	<r.>	<U1E59>	LATIN SMALL LETTER R WITH DOT ABOVE
4710	<R-.>	<U1E5A>	LATIN CAPITAL LETTER R WITH DOT BELOW
4711	<r-.>	<U1E5B>	LATIN SMALL LETTER R WITH DOT BELOW
4712	<R--.>	<U1E5C>	LATIN CAPITAL LETTER R WITH DOT BELOW AND MACRON
4713	<r--.>	<U1E5D>	LATIN SMALL LETTER R WITH DOT BELOW AND MACRON
4714	<R_>	<U1E5E>	LATIN CAPITAL LETTER R WITH LINE BELOW
4715	<r_>	<U1E5F>	LATIN SMALL LETTER R WITH LINE BELOW
4716	<S.>	<U1E60>	LATIN CAPITAL LETTER S WITH DOT ABOVE
4717	<s.>	<U1E61>	LATIN SMALL LETTER S WITH DOT ABOVE
4718	<S-.>	<U1E62>	LATIN CAPITAL LETTER S WITH DOT BELOW
4719	<s-.>	<U1E63>	LATIN SMALL LETTER S WITH DOT BELOW
4720	<S'>	<U1E64>	LATIN CAPITAL LETTER S WITH ACUTE AND DOT ABOVE
4721	<s'>	<U1E65>	LATIN SMALL LETTER S WITH ACUTE AND DOT ABOVE
4722	<S<.>	<U1E66>	LATIN CAPITAL LETTER S WITH CARON AND DOT ABOVE
4723	<s<.>	<U1E67>	LATIN SMALL LETTER S WITH CARON AND DOT ABOVE
4724	<S.-.>	<U1E68>	LATIN CAPITAL LETTER S WITH DOT BELOW AND DOT ABOVE
4725	<s.-.>	<U1E69>	LATIN SMALL LETTER S WITH DOT BELOW AND DOT ABOVE
4726	<T.>	<U1E6A>	LATIN CAPITAL LETTER T WITH DOT ABOVE
4727	<t.>	<U1E6B>	LATIN SMALL LETTER T WITH DOT ABOVE
4728	<T-.>	<U1E6C>	LATIN CAPITAL LETTER T WITH DOT BELOW
4729	<t-.>	<U1E6D>	LATIN SMALL LETTER T WITH DOT BELOW
4730	<T_>	<U1E6E>	LATIN CAPITAL LETTER T WITH LINE BELOW
4731	<t_>	<U1E6F>	LATIN SMALL LETTER T WITH LINE BELOW
4732	<T-/>	<U1E70>	LATIN CAPITAL LETTER T WITH CIRCUMFLEX BELOW
4733	<t-/>	<U1E71>	LATIN SMALL LETTER T WITH CIRCUMFLEX BELOW
4734	<U--.>	<U1E72>	LATIN CAPITAL LETTER U WITH DIAERESIS BELOW
4735	<u--.>	<U1E73>	LATIN SMALL LETTER U WITH DIAERESIS BELOW
4736	<U-?>	<U1E74>	LATIN CAPITAL LETTER U WITH TILDE BELOW
4737	<u-?>	<U1E75>	LATIN SMALL LETTER U WITH TILDE BELOW
4738	<U-/>	<U1E76>	LATIN CAPITAL LETTER U WITH CIRCUMFLEX BELOW
4739	<u-/>	<U1E77>	LATIN SMALL LETTER U WITH CIRCUMFLEX BELOW
4740	<U?'>	<U1E78>	LATIN CAPITAL LETTER U WITH TILDE AND ACUTE
4741	<u?'>	<U1E79>	LATIN SMALL LETTER U WITH TILDE AND ACUTE
4742	<U-:>	<U1E7A>	LATIN CAPITAL LETTER U WITH MACRON AND DIAERESIS
4743	<u-:>	<U1E7B>	LATIN SMALL LETTER U WITH MACRON AND DIAERESIS
4744	<V?>	<U1E7C>	LATIN CAPITAL LETTER V WITH TILDE
4745	<v?>	<U1E7D>	LATIN SMALL LETTER V WITH TILDE
4746	<V-.>	<U1E7E>	LATIN CAPITAL LETTER V WITH DOT BELOW
4747	<v-.>	<U1E7F>	LATIN SMALL LETTER V WITH DOT BELOW
4748	<W!>	<U1E80>	LATIN CAPITAL LETTER W WITH GRAVE
4749	<w!>	<U1E81>	LATIN SMALL LETTER W WITH GRAVE
4750	<W'>	<U1E82>	LATIN CAPITAL LETTER W WITH ACUTE
4751	<w'>	<U1E83>	LATIN SMALL LETTER W WITH ACUTE
4752	<W.>	<U1E84>	LATIN CAPITAL LETTER W WITH DIAERESIS
4753	<w.>	<U1E85>	LATIN SMALL LETTER W WITH DIAERESIS
4754	<W.>	<U1E86>	LATIN CAPITAL LETTER W WITH DOT ABOVE
4755	<w.>	<U1E87>	LATIN SMALL LETTER W WITH DOT ABOVE
4756	<W-.>	<U1E88>	LATIN CAPITAL LETTER W WITH DOT BELOW
4757	<w-.>	<U1E89>	LATIN SMALL LETTER W WITH DOT BELOW
4758	<X.>	<U1E8A>	LATIN CAPITAL LETTER X WITH DOT ABOVE
4759	<x.>	<U1E8B>	LATIN SMALL LETTER X WITH DOT ABOVE

4760	<x:>	<U1E8C>	LATIN CAPITAL LETTER X WITH DIAERESIS
4761	<x:>	<U1E8D>	LATIN SMALL LETTER X WITH DIAERESIS
4762	<y.>	<U1E8E>	LATIN CAPITAL LETTER Y WITH DOT ABOVE
4763	<y.>	<U1E8F>	LATIN SMALL LETTER Y WITH DOT ABOVE
4764	<z/>>	<U1E90>	LATIN CAPITAL LETTER Z WITH CIRCUMFLEX
4765	<z/>>	<U1E91>	LATIN SMALL LETTER Z WITH CIRCUMFLEX
4766	<z-.>	<U1E92>	LATIN CAPITAL LETTER Z WITH DOT BELOW
4767	<z-.>	<U1E93>	LATIN SMALL LETTER Z WITH DOT BELOW
4768	<z_>	<U1E94>	LATIN CAPITAL LETTER Z WITH LINE BELOW
4769	<z_>	<U1E95>	LATIN SMALL LETTER Z WITH LINE BELOW
4770	<a-.>	<U1EA0>	LATIN CAPITAL LETTER A WITH DOT BELOW
4771	<a-.>	<U1EA1>	LATIN SMALL LETTER A WITH DOT BELOW
4772	<a2>	<U1EA2>	LATIN CAPITAL LETTER A WITH HOOK ABOVE
4773	<a2>	<U1EA3>	LATIN SMALL LETTER A WITH HOOK ABOVE
4774	<a/>'>	<U1EA4>	LATIN CAPITAL LETTER A WITH CIRCUMFLEX AND ACUTE
4775	<a/>'>	<U1EA5>	LATIN SMALL LETTER A WITH CIRCUMFLEX AND ACUTE
4776	<a/>!>	<U1EA6>	LATIN CAPITAL LETTER A WITH CIRCUMFLEX AND GRAVE
4777	<a/>!>	<U1EA7>	LATIN SMALL LETTER A WITH CIRCUMFLEX AND GRAVE
4778	<a/>2>	<U1EA8>	LATIN CAPITAL LETTER A WITH CIRCUMFLEX AND HOOK ABOVE
4779	<a/>2>	<U1EA9>	LATIN SMALL LETTER A WITH CIRCUMFLEX AND HOOK ABOVE
4780	<a/>?>	<U1EAA>	LATIN CAPITAL LETTER A WITH CIRCUMFLEX AND TILDE
4781	<a/>?>	<U1EAB>	LATIN SMALL LETTER A WITH CIRCUMFLEX AND TILDE
4782	<a/>-.>	<U1EAC>	LATIN CAPITAL LETTER A WITH CIRCUMFLEX AND DOT BELOW
4783	<a/>-.>	<U1EAD>	LATIN SMALL LETTER A WITH CIRCUMFLEX AND DOT BELOW
4784	<a(')>	<U1EAE>	LATIN CAPITAL LETTER A WITH BREVE AND ACUTE
4785	<a(')>	<U1EAF>	LATIN SMALL LETTER A WITH BREVE AND ACUTE
4786	<a(!)>	<U1EB0>	LATIN CAPITAL LETTER A WITH BREVE AND GRAVE
4787	<a(!)>	<U1EB1>	LATIN SMALL LETTER A WITH BREVE AND GRAVE
4788	<a(2)>	<U1EB2>	LATIN CAPITAL LETTER A WITH BREVE AND HOOK ABOVE
4789	<a(2)>	<U1EB3>	LATIN SMALL LETTER A WITH BREVE AND HOOK ABOVE
4790	<a(?)>	<U1EB4>	LATIN CAPITAL LETTER A WITH BREVE AND TILDE
4791	<a(?)>	<U1EB5>	LATIN SMALL LETTER A WITH BREVE AND TILDE
4792	<a(-.>	<U1EB6>	LATIN CAPITAL LETTER A WITH BREVE AND DOT BELOW
4793	<a(-.>	<U1EB7>	LATIN SMALL LETTER A WITH BREVE AND DOT BELOW
4794	<e-.>	<U1EB8>	LATIN CAPITAL LETTER E WITH DOT BELOW
4795	<e-.>	<U1EB9>	LATIN SMALL LETTER E WITH DOT BELOW
4796	<e2>	<U1EBA>	LATIN CAPITAL LETTER E WITH HOOK ABOVE
4797	<e2>	<U1EBB>	LATIN SMALL LETTER E WITH HOOK ABOVE
4798	<e?>	<U1EBC>	LATIN CAPITAL LETTER E WITH TILDE
4799	<e?>	<U1EBD>	LATIN SMALL LETTER E WITH TILDE
4800	<e/>'>	<U1EBE>	LATIN CAPITAL LETTER E WITH CIRCUMFLEX AND ACUTE
4801	<e/>'>	<U1EBF>	LATIN SMALL LETTER E WITH CIRCUMFLEX AND ACUTE
4802	<e/>!>	<U1EC0>	LATIN CAPITAL LETTER E WITH CIRCUMFLEX AND GRAVE
4803	<e/>!>	<U1EC1>	LATIN SMALL LETTER E WITH CIRCUMFLEX AND GRAVE
4804	<e/>2>	<U1EC2>	LATIN CAPITAL LETTER E WITH CIRCUMFLEX AND HOOK ABOVE
4805	<e/>2>	<U1EC3>	LATIN SMALL LETTER E WITH CIRCUMFLEX AND HOOK ABOVE
4806	<e/>?>	<U1EC4>	LATIN CAPITAL LETTER E WITH CIRCUMFLEX AND TILDE
4807	<e/>?>	<U1EC5>	LATIN SMALL LETTER E WITH CIRCUMFLEX AND TILDE
4808	<e/>-.>	<U1EC6>	LATIN CAPITAL LETTER E WITH CIRCUMFLEX AND DOT BELOW
4809	<e/>-.>	<U1EC7>	LATIN SMALL LETTER E WITH CIRCUMFLEX AND DOT BELOW
4810	<i2>	<U1EC8>	LATIN CAPITAL LETTER I WITH HOOK ABOVE
4811	<i2>	<U1EC9>	LATIN SMALL LETTER I WITH HOOK ABOVE
4812	<i-.>	<U1ECA>	LATIN CAPITAL LETTER I WITH DOT BELOW
4813	<i-.>	<U1ECB>	LATIN SMALL LETTER I WITH DOT BELOW
4814	<o-.>	<U1ECC>	LATIN CAPITAL LETTER O WITH DOT BELOW
4815	<o-.>	<U1ECD>	LATIN SMALL LETTER O WITH DOT BELOW
4816	<o2>	<U1ECE>	LATIN CAPITAL LETTER O WITH HOOK ABOVE
4817	<o2>	<U1ECF>	LATIN SMALL LETTER O WITH HOOK ABOVE
4818	<o/>'>	<U1ED0>	LATIN CAPITAL LETTER O WITH CIRCUMFLEX AND ACUTE
4819	<o/>'>	<U1ED1>	LATIN SMALL LETTER O WITH CIRCUMFLEX AND ACUTE
4820	<o/>!>	<U1ED2>	LATIN CAPITAL LETTER O WITH CIRCUMFLEX AND GRAVE
4821	<o/>!>	<U1ED3>	LATIN SMALL LETTER O WITH CIRCUMFLEX AND GRAVE
4822	<o/>2>	<U1ED4>	LATIN CAPITAL LETTER O WITH CIRCUMFLEX AND HOOK ABOVE
4823	<o/>2>	<U1ED5>	LATIN SMALL LETTER O WITH CIRCUMFLEX AND HOOK ABOVE
4824	<o/>?>	<U1ED6>	LATIN CAPITAL LETTER O WITH CIRCUMFLEX AND TILDE
4825	<o/>?>	<U1ED7>	LATIN SMALL LETTER O WITH CIRCUMFLEX AND TILDE
4826	<o/>-.>	<U1ED8>	LATIN CAPITAL LETTER O WITH CIRCUMFLEX AND DOT BELOW
4827	<o/>-.>	<U1ED9>	LATIN SMALL LETTER O WITH CIRCUMFLEX AND DOT BELOW
4828	<o9'>	<U1EDA>	LATIN CAPITAL LETTER O WITH HORN AND ACUTE
4829	<o9'>	<U1EDB>	LATIN SMALL LETTER O WITH HORN AND ACUTE
4830	<o9!>	<U1EDC>	LATIN CAPITAL LETTER O WITH HORN AND GRAVE
4831	<o9!>	<U1EDD>	LATIN SMALL LETTER O WITH HORN AND GRAVE
4832	<o92>	<U1EDE>	LATIN CAPITAL LETTER O WITH HORN AND HOOK ABOVE
4833	<o92>	<U1EDF>	LATIN SMALL LETTER O WITH HORN AND HOOK ABOVE
4834	<o9?>	<U1EE0>	LATIN CAPITAL LETTER O WITH HORN AND TILDE
4835	<o9?>	<U1EE1>	LATIN SMALL LETTER O WITH HORN AND TILDE
4836	<o9-.>	<U1EE2>	LATIN CAPITAL LETTER O WITH HORN AND DOT BELOW
4837	<o9-.>	<U1EE3>	LATIN SMALL LETTER O WITH HORN AND DOT BELOW
4838	<u-.>	<U1EE4>	LATIN CAPITAL LETTER U WITH DOT BELOW
4839	<u-.>	<U1EE5>	LATIN SMALL LETTER U WITH DOT BELOW
4840	<u2>	<U1EE6>	LATIN CAPITAL LETTER U WITH HOOK ABOVE
4841	<u2>	<U1EE7>	LATIN SMALL LETTER U WITH HOOK ABOVE
4842	<u9'>	<U1EE8>	LATIN CAPITAL LETTER U WITH HORN AND ACUTE
4843	<u9'>	<U1EE9>	LATIN SMALL LETTER U WITH HORN AND ACUTE
4844	<u9!>	<U1EEA>	LATIN CAPITAL LETTER U WITH HORN AND GRAVE
4845	<u9!>	<U1EEB>	LATIN SMALL LETTER U WITH HORN AND GRAVE
4846	<u92>	<U1EEC>	LATIN CAPITAL LETTER U WITH HORN AND HOOK ABOVE



4847	<u92>	<U1EED>	LATIN SMALL LETTER U WITH HORN AND HOOK ABOVE
4848	<U9?>	<U1EEE>	LATIN CAPITAL LETTER U WITH HORN AND TILDE
4849	<u9?>	<U1EEF>	LATIN SMALL LETTER U WITH HORN AND TILDE
4850	<U9-.>	<U1EF0>	LATIN CAPITAL LETTER U WITH HORN AND DOT BELOW
4851	<u9-.>	<U1EF1>	LATIN SMALL LETTER U WITH HORN AND DOT BELOW
4852	<Y!>	<U1EF2>	LATIN CAPITAL LETTER Y WITH GRAVE
4853	<y!>	<U1EF3>	LATIN SMALL LETTER Y WITH GRAVE
4854	<Y-.>	<U1EF4>	LATIN CAPITAL LETTER Y WITH DOT BELOW
4855	<y-.>	<U1EF5>	LATIN SMALL LETTER Y WITH DOT BELOW
4856	<Y2>	<U1EF6>	LATIN CAPITAL LETTER Y WITH HOOK ABOVE
4857	<y2>	<U1EF7>	LATIN SMALL LETTER Y WITH HOOK ABOVE
4858	<Y?>	<U1EF8>	LATIN CAPITAL LETTER Y WITH TILDE
4859	<y?>	<U1EF9>	LATIN SMALL LETTER Y WITH TILDE
4860	<a*,>	<U1F00>	GREEK SMALL LETTER ALPHA WITH PSILI
4861	<a*;>	<U1F01>	GREEK SMALL LETTER ALPHA WITH DASIA
4862	<a*;!>	<U1F02>	GREEK SMALL LETTER ALPHA WITH PSILI AND VARIA
4863	<a*;!>	<U1F03>	GREEK SMALL LETTER ALPHA WITH DASIA AND VARIA
4864	<a*,'>	<U1F04>	GREEK SMALL LETTER ALPHA WITH PSILI AND OXIA
4865	<a*,'>	<U1F05>	GREEK SMALL LETTER ALPHA WITH DASIA AND OXIA
4866	<a*,>?	<U1F06>	GREEK SMALL LETTER ALPHA WITH PSILI AND PERISPOMENI
4867	<a*;>?	<U1F07>	GREEK SMALL LETTER ALPHA WITH DASIA AND PERISPOMENI
4868	<A*,>	<U1F08>	GREEK CAPITAL LETTER ALPHA WITH PSILI
4869	<A*,>	<U1F09>	GREEK CAPITAL LETTER ALPHA WITH DASIA
4870	<A*;!>	<U1F0A>	GREEK CAPITAL LETTER ALPHA WITH PSILI AND VARIA
4871	<A*;!>	<U1F0B>	GREEK CAPITAL LETTER ALPHA WITH DASIA AND VARIA
4872	<A*,'>	<U1F0C>	GREEK CAPITAL LETTER ALPHA WITH PSILI AND OXIA
4873	<A*,'>	<U1F0D>	GREEK CAPITAL LETTER ALPHA WITH DASIA AND OXIA
4874	<A*,>?	<U1F0E>	GREEK CAPITAL LETTER ALPHA WITH PSILI AND PERISPOMENI
4875	<A*,>?	<U1F0F>	GREEK CAPITAL LETTER ALPHA WITH DASIA AND PERISPOMENI
4876	<e*,>	<U1F10>	GREEK SMALL LETTER EPSILON WITH PSILI
4877	<e*;>	<U1F11>	GREEK SMALL LETTER EPSILON WITH DASIA
4878	<e*;!>	<U1F12>	GREEK SMALL LETTER EPSILON WITH PSILI AND VARIA
4879	<e*;!>	<U1F13>	GREEK SMALL LETTER EPSILON WITH DASIA AND VARIA
4880	<e*,'>	<U1F14>	GREEK SMALL LETTER EPSILON WITH PSILI AND OXIA
4881	<e*,'>	<U1F15>	GREEK SMALL LETTER EPSILON WITH DASIA AND OXIA
4882	<E*,>	<U1F18>	GREEK CAPITAL LETTER EPSILON WITH PSILI
4883	<E*,>	<U1F19>	GREEK CAPITAL LETTER EPSILON WITH DASIA
4884	<E*;!>	<U1F1A>	GREEK CAPITAL LETTER EPSILON WITH PSILI AND VARIA
4885	<E*;!>	<U1F1B>	GREEK CAPITAL LETTER EPSILON WITH DASIA AND VARIA
4886	<E*,'>	<U1F1C>	GREEK CAPITAL LETTER EPSILON WITH PSILI AND OXIA
4887	<E*,'>	<U1F1D>	GREEK CAPITAL LETTER EPSILON WITH DASIA AND OXIA
4888	<y*,>	<U1F20>	GREEK SMALL LETTER ETA WITH PSILI
4889	<y*;>	<U1F21>	GREEK SMALL LETTER ETA WITH DASIA
4890	<y*;!>	<U1F22>	GREEK SMALL LETTER ETA WITH PSILI AND VARIA
4891	<y*;!>	<U1F23>	GREEK SMALL LETTER ETA WITH DASIA AND VARIA
4892	<y*,'>	<U1F24>	GREEK SMALL LETTER ETA WITH PSILI AND OXIA
4893	<y*,'>	<U1F25>	GREEK SMALL LETTER ETA WITH DASIA AND OXIA
4894	<y*,>?	<U1F26>	GREEK SMALL LETTER ETA WITH PSILI AND PERISPOMENI
4895	<y*;>?	<U1F27>	GREEK SMALL LETTER ETA WITH DASIA AND PERISPOMENI
4896	<Y*,>	<U1F28>	GREEK CAPITAL LETTER ETA WITH PSILI
4897	<Y*,>	<U1F29>	GREEK CAPITAL LETTER ETA WITH DASIA
4898	<Y*;!>	<U1F2A>	GREEK CAPITAL LETTER ETA WITH PSILI AND VARIA
4899	<Y*;!>	<U1F2B>	GREEK CAPITAL LETTER ETA WITH DASIA AND VARIA
4900	<Y*,'>	<U1F2C>	GREEK CAPITAL LETTER ETA WITH PSILI AND OXIA
4901	<Y*,'>	<U1F2D>	GREEK CAPITAL LETTER ETA WITH DASIA AND OXIA
4902	<Y*,>?	<U1F2E>	GREEK CAPITAL LETTER ETA WITH PSILI AND PERISPOMENI
4903	<Y*;>?	<U1F2F>	GREEK CAPITAL LETTER ETA WITH DASIA AND PERISPOMENI
4904	<i*,>	<U1F30>	GREEK SMALL LETTER IOTA WITH PSILI
4905	<i*;>	<U1F31>	GREEK SMALL LETTER IOTA WITH DASIA
4906	<i*;!>	<U1F32>	GREEK SMALL LETTER IOTA WITH PSILI AND VARIA
4907	<i*;!>	<U1F33>	GREEK SMALL LETTER IOTA WITH DASIA AND VARIA
4908	<i*,'>	<U1F34>	GREEK SMALL LETTER IOTA WITH PSILI AND OXIA
4909	<i*,'>	<U1F35>	GREEK SMALL LETTER IOTA WITH DASIA AND OXIA
4910	<i*,>?	<U1F36>	GREEK SMALL LETTER IOTA WITH PSILI AND PERISPOMENI
4911	<i*;>?	<U1F37>	GREEK SMALL LETTER IOTA WITH DASIA AND PERISPOMENI
4912	<I*,>	<U1F38>	GREEK CAPITAL LETTER IOTA WITH PSILI
4913	<I*;>	<U1F39>	GREEK CAPITAL LETTER IOTA WITH DASIA
4914	<I*;!>	<U1F3A>	GREEK CAPITAL LETTER IOTA WITH PSILI AND VARIA
4915	<I*;!>	<U1F3B>	GREEK CAPITAL LETTER IOTA WITH DASIA AND VARIA
4916	<I*,'>	<U1F3C>	GREEK CAPITAL LETTER IOTA WITH PSILI AND OXIA
4917	<I*,'>	<U1F3D>	GREEK CAPITAL LETTER IOTA WITH DASIA AND OXIA
4918	<I*,>?	<U1F3E>	GREEK CAPITAL LETTER IOTA WITH PSILI AND PERISPOMENI
4919	<I*;>?	<U1F3F>	GREEK CAPITAL LETTER IOTA WITH DASIA AND PERISPOMENI
4920	<o*,>	<U1F40>	GREEK SMALL LETTER OMICRON WITH PSILI
4921	<o*;>	<U1F41>	GREEK SMALL LETTER OMICRON WITH DASIA
4922	<o*;!>	<U1F42>	GREEK SMALL LETTER OMICRON WITH PSILI AND VARIA
4923	<o*;!>	<U1F43>	GREEK SMALL LETTER OMICRON WITH DASIA AND VARIA
4924	<o*,'>	<U1F44>	GREEK SMALL LETTER OMICRON WITH PSILI AND OXIA
4925	<o*,'>	<U1F45>	GREEK SMALL LETTER OMICRON WITH DASIA AND OXIA
4926	<O*,>	<U1F48>	GREEK CAPITAL LETTER OMICRON WITH PSILI
4927	<O*,>	<U1F49>	GREEK CAPITAL LETTER OMICRON WITH DASIA
4928	<O*;!>	<U1F4A>	GREEK CAPITAL LETTER OMICRON WITH PSILI AND VARIA
4929	<O*;!>	<U1F4B>	GREEK CAPITAL LETTER OMICRON WITH DASIA AND VARIA
4930	<O*,'>	<U1F4C>	GREEK CAPITAL LETTER OMICRON WITH PSILI AND OXIA
4931	<O*,'>	<U1F4D>	GREEK CAPITAL LETTER OMICRON WITH DASIA AND OXIA
4932	<u*,>	<U1F50>	GREEK SMALL LETTER UPSILON WITH PSILI
4933	<u*;>	<U1F51>	GREEK SMALL LETTER UPSILON WITH DASIA
4934	<u*;!>	<U1F52>	GREEK SMALL LETTER UPSILON WITH PSILI AND VARIA
4935	<u*;!>	<U1F53>	GREEK SMALL LETTER UPSILON WITH DASIA AND VARIA

4936	<u*, ' >	<U1F54>	GREEK SMALL LETTER UPSILON WITH PSILI AND OXIA
4937	<u*, ' >	<U1F55>	GREEK SMALL LETTER UPSILON WITH DASIA AND OXIA
4938	<u*, ? >	<U1F56>	GREEK SMALL LETTER UPSILON WITH PSILI AND PERISPOMENI
4939	<u*, ? >	<U1F57>	GREEK SMALL LETTER UPSILON WITH DASIA AND PERISPOMENI
4940	<U* ; >	<U1F59>	GREEK CAPITAL LETTER UPSILON WITH DASIA
4941	<U* ; ! >	<U1F5B>	GREEK CAPITAL LETTER UPSILON WITH DASIA AND VARIA
4942	<U* ; ' >	<U1F5D>	GREEK CAPITAL LETTER UPSILON WITH DASIA AND OXIA
4943	<U* ; ? >	<U1F5F>	GREEK CAPITAL LETTER UPSILON WITH DASIA AND PERISPOMENI
4944	<w* , >	<U1F60>	GREEK SMALL LETTER OMEGA WITH PSILI
4945	<w* ; >	<U1F61>	GREEK SMALL LETTER OMEGA WITH DASIA
4946	<w* , ! >	<U1F62>	GREEK SMALL LETTER OMEGA WITH PSILI AND VARIA
4947	<w* ; ! >	<U1F63>	GREEK SMALL LETTER OMEGA WITH DASIA AND VARIA
4948	<w* , ' >	<U1F64>	GREEK SMALL LETTER OMEGA WITH PSILI AND OXIA
4949	<w* ; ' >	<U1F65>	GREEK SMALL LETTER OMEGA WITH DASIA AND OXIA
4950	<w* , ? >	<U1F66>	GREEK SMALL LETTER OMEGA WITH PSILI AND PERISPOMENI
4951	<w* ; ? >	<U1F67>	GREEK SMALL LETTER OMEGA WITH DASIA AND PERISPOMENI
4952	<W* , >	<U1F68>	GREEK CAPITAL LETTER OMEGA WITH PSILI
4953	<W* ; >	<U1F69>	GREEK CAPITAL LETTER OMEGA WITH DASIA
4954	<W* , ! >	<U1F6A>	GREEK CAPITAL LETTER OMEGA WITH PSILI AND VARIA
4955	<W* ; ! >	<U1F6B>	GREEK CAPITAL LETTER OMEGA WITH DASIA AND VARIA
4956	<W* , ' >	<U1F6C>	GREEK CAPITAL LETTER OMEGA WITH PSILI AND OXIA
4957	<W* ; ' >	<U1F6D>	GREEK CAPITAL LETTER OMEGA WITH DASIA AND OXIA
4958	<W* , ? >	<U1F6E>	GREEK CAPITAL LETTER OMEGA WITH PSILI AND PERISPOMENI
4959	<W* ; ? >	<U1F6F>	GREEK CAPITAL LETTER OMEGA WITH DASIA AND PERISPOMENI
4960	<a* ! >	<U1F70>	GREEK SMALL LETTER ALPHA WITH VARIA
4961	<a* * >	<U1F71>	GREEK SMALL LETTER ALPHA WITH OXIA
4962	<e* ! >	<U1F72>	GREEK SMALL LETTER EPSILON WITH VARIA
4963	<e* * >	<U1F73>	GREEK SMALL LETTER EPSILON WITH OXIA
4964	<y* ! >	<U1F74>	GREEK SMALL LETTER ETA WITH VARIA
4965	<y* * >	<U1F75>	GREEK SMALL LETTER ETA WITH OXIA
4966	<i* ! >	<U1F76>	GREEK SMALL LETTER IOTA WITH VARIA
4967	<i* * >	<U1F77>	GREEK SMALL LETTER IOTA WITH OXIA
4968	<o* ! >	<U1F78>	GREEK SMALL LETTER OMICRON WITH VARIA
4969	<o* * >	<U1F79>	GREEK SMALL LETTER OMICRON WITH OXIA
4970	<u* ! >	<U1F7A>	GREEK SMALL LETTER UPSILON WITH VARIA
4971	<u* * >	<U1F7B>	GREEK SMALL LETTER UPSILON WITH OXIA
4972	<w* ! >	<U1F7C>	GREEK SMALL LETTER OMEGA WITH VARIA
4973	<w* * >	<U1F7D>	GREEK SMALL LETTER OMEGA WITH OXIA
4974	<a* , j >	<U1F80>	GREEK SMALL LETTER ALPHA WITH PSILI AND YPOGEGRAMMENI
4975	<a* ; j >	<U1F81>	GREEK SMALL LETTER ALPHA WITH DASIA AND YPOGEGRAMMENI
4976	<a* , ! j >	<U1F82>	GREEK SMALL LETTER ALPHA WITH PSILI AND VARIA AND YPOGEGRAMMENI
4977	<a* ; ! j >	<U1F83>	GREEK SMALL LETTER ALPHA WITH DASIA AND VARIA AND YPOGEGRAMMENI
4978	<a* , ' j >	<U1F84>	GREEK SMALL LETTER ALPHA WITH PSILI AND OXIA AND YPOGEGRAMMENI
4979	<a* ; ' j >	<U1F85>	GREEK SMALL LETTER ALPHA WITH DASIA AND OXIA AND YPOGEGRAMMENI
4980	<a* , ? j >	<U1F86>	GREEK SMALL LETTER ALPHA WITH PSILI AND PERISPOMENI AND YPOGEGRAMMENI
4981	<a* ; ? j >	<U1F87>	GREEK SMALL LETTER ALPHA WITH DASIA AND PERISPOMENI AND YPOGEGRAMMENI
4982	<A* , J >	<U1F88>	GREEK CAPITAL LETTER ALPHA WITH PSILI AND PROSGEGRAMMENI
4983	<A* ; J >	<U1F89>	GREEK CAPITAL LETTER ALPHA WITH DASIA AND PROSGEGRAMMENI
4984	<A* , ! J >	<U1F8A>	GREEK CAPITAL LETTER ALPHA WITH PSILI AND VARIA AND PROSGEGRAMMENI
4985	<A* ; ! J >	<U1F8B>	GREEK CAPITAL LETTER ALPHA WITH DASIA AND VARIA AND PROSGEGRAMMENI
4986	<A* , ' J >	<U1F8C>	GREEK CAPITAL LETTER ALPHA WITH PSILI AND OXIA AND PROSGEGRAMMENI
4987	<A* ; ' J >	<U1F8D>	GREEK CAPITAL LETTER ALPHA WITH DASIA AND OXIA AND PROSGEGRAMMENI
4988	<A* , ? J >	<U1F8E>	GREEK CAPITAL LETTER ALPHA WITH PSILI AND PERISPOMENI AND PROSGEGRAMMENI
4989	<A* ; ? J >	<U1F8F>	GREEK CAPITAL LETTER ALPHA WITH DASIA AND PERISPOMENI AND PROSGEGRAMMENI
4990	<y* , j >	<U1F90>	GREEK SMALL LETTER ETA WITH PSILI AND YPOGEGRAMMENI
4991	<y* ; j >	<U1F91>	GREEK SMALL LETTER ETA WITH DASIA AND YPOGEGRAMMENI
4992	<y* , ! j >	<U1F92>	GREEK SMALL LETTER ETA WITH PSILI AND VARIA AND YPOGEGRAMMENI
4993	<y* ; ! j >	<U1F93>	GREEK SMALL LETTER ETA WITH DASIA AND VARIA AND YPOGEGRAMMENI
4994	<y* , ' j >	<U1F94>	GREEK SMALL LETTER ETA WITH PSILI AND OXIA AND YPOGEGRAMMENI
4995	<y* ; ' j >	<U1F95>	GREEK SMALL LETTER ETA WITH DASIA AND OXIA AND YPOGEGRAMMENI
4996	<y* , ? j >	<U1F96>	GREEK SMALL LETTER ETA WITH PSILI AND PERISPOMENI AND YPOGEGRAMMENI
4997	<y* ; ? j >	<U1F97>	GREEK SMALL LETTER ETA WITH DASIA AND PERISPOMENI AND YPOGEGRAMMENI
4998	<Y* , J >	<U1F98>	GREEK CAPITAL LETTER ETA WITH PSILI AND PROSGEGRAMMENI
4999	<Y* ; J >	<U1F99>	GREEK CAPITAL LETTER ETA WITH DASIA AND PROSGEGRAMMENI
5000	<Y* , ! J >	<U1F9A>	GREEK CAPITAL LETTER ETA WITH PSILI AND VARIA AND PROSGEGRAMMENI
5001	<Y* ; ! J >	<U1F9B>	GREEK CAPITAL LETTER ETA WITH DASIA AND VARIA AND PROSGEGRAMMENI
5002	<Y* , ' J >	<U1F9C>	GREEK CAPITAL LETTER ETA WITH PSILI AND OXIA AND PROSGEGRAMMENI
5003	<Y* ; ' J >	<U1F9D>	GREEK CAPITAL LETTER ETA WITH DASIA AND OXIA AND PROSGEGRAMMENI
5004	<Y* , ? J >	<U1F9E>	GREEK CAPITAL LETTER ETA WITH PSILI AND PERISPOMENI AND PROSGEGRAMMENI
5005	<Y* ; ? J >	<U1F9F>	GREEK CAPITAL LETTER ETA WITH DASIA AND PERISPOMENI AND PROSGEGRAMMENI
5006	<w* , j >	<U1FA0>	GREEK SMALL LETTER OMEGA WITH PSILI AND YPOGEGRAMMENI
5007	<w* ; j >	<U1FA1>	GREEK SMALL LETTER OMEGA WITH DASIA AND YPOGEGRAMMENI
5008	<w* , ! j >	<U1FA2>	GREEK SMALL LETTER OMEGA WITH PSILI AND VARIA AND YPOGEGRAMMENI
5009	<w* ; ! j >	<U1FA3>	GREEK SMALL LETTER OMEGA WITH DASIA AND VARIA AND YPOGEGRAMMENI
5010	<w* , ' j >	<U1FA4>	GREEK SMALL LETTER OMEGA WITH PSILI AND OXIA AND YPOGEGRAMMENI
5011	<w* ; ' j >	<U1FA5>	GREEK SMALL LETTER OMEGA WITH DASIA AND OXIA AND YPOGEGRAMMENI
5012	<w* , ? j >	<U1FA6>	GREEK SMALL LETTER OMEGA WITH PSILI AND PERISPOMENI AND YPOGEGRAMMENI
5013	<w* ; ? j >	<U1FA7>	GREEK SMALL LETTER OMEGA WITH DASIA AND PERISPOMENI AND YPOGEGRAMMENI
5014	<W* , J >	<U1FA8>	GREEK CAPITAL LETTER OMEGA WITH PSILI AND PROSGEGRAMMENI
5015	<W* ; J >	<U1FA9>	GREEK CAPITAL LETTER OMEGA WITH DASIA AND PROSGEGRAMMENI
5016	<W* , ! J >	<U1FAA>	GREEK CAPITAL LETTER OMEGA WITH PSILI AND VARIA AND PROSGEGRAMMENI
5017	<W* ; ! J >	<U1FAB>	GREEK CAPITAL LETTER OMEGA WITH DASIA AND VARIA AND PROSGEGRAMMENI
5018	<W* , ' J >	<U1FAC>	GREEK CAPITAL LETTER OMEGA WITH PSILI AND OXIA AND PROSGEGRAMMENI
5019	<W* ; ' J >	<U1FAD>	GREEK CAPITAL LETTER OMEGA WITH DASIA AND OXIA AND PROSGEGRAMMENI
5020	<W* , ? J >	<U1FAE>	GREEK CAPITAL LETTER OMEGA WITH PSILI AND PERISPOMENI AND PROSGEGRAMMENI
5021	<W* ; ? J >	<U1FAF>	GREEK CAPITAL LETTER OMEGA WITH DASIA AND PERISPOMENI AND PROSGEGRAMMENI
5022	<a* ( >	<U1FB0>	GREEK SMALL LETTER ALPHA WITH VRACHY

5023	<a*->	<U1FB1>	GREEK SMALL LETTER ALPHA WITH MACRON
5024	<a*!j>	<U1FB2>	GREEK SMALL LETTER ALPHA WITH VARIA AND YPOGEGRAMMENI
5025	<a*j>	<U1FB3>	GREEK SMALL LETTER ALPHA WITH YPOGEGRAMMENI
5026	<a*~j>	<U1FB4>	GREEK SMALL LETTER ALPHA WITH OXIA AND YPOGEGRAMMENI
5027	<a*?>	<U1FB6>	GREEK SMALL LETTER ALPHA WITH PERISPOMENI
5028	<a*?j>	<U1FB7>	GREEK SMALL LETTER ALPHA WITH PERISPOMENI AND YPOGEGRAMMENI
5029	<A*(>	<U1FB8>	GREEK CAPITAL LETTER ALPHA WITH VRACHY
5030	<A*->	<U1FB9>	GREEK CAPITAL LETTER ALPHA WITH MACRON
5031	<A*!>	<U1FBA>	GREEK CAPITAL LETTER ALPHA WITH VARIA
5032	<A*~>	<U1FBB>	GREEK CAPITAL LETTER ALPHA WITH OXIA
5033	<A*J>	<U1FBC>	GREEK CAPITAL LETTER ALPHA WITH PROSGEGRAMMENI
5034	<)*>	<U1FBD>	GREEK KORONIS
5035	<J3>	<U1FBE>	GREEK PROSGEGRAMMENI
5036	<, ,>	<U1FBF>	GREEK PSILI
5037	<?*>	<U1FC0>	GREEK PERISPOMENI
5038	<?:>	<U1FC1>	GREEK DIALYTIKA AND PERISPOMENI
5039	<y*!j>	<U1FC2>	GREEK SMALL LETTER ETA WITH VARIA AND YPOGEGRAMMENI
5040	<y*j>	<U1FC3>	GREEK SMALL LETTER ETA WITH YPOGEGRAMMENI
5041	<y*~j>	<U1FC4>	GREEK SMALL LETTER ETA WITH OXIA AND YPOGEGRAMMENI
5042	<y*?>	<U1FC6>	GREEK SMALL LETTER ETA WITH PERISPOMENI
5043	<y*?j>	<U1FC7>	GREEK SMALL LETTER ETA WITH PERISPOMENI AND YPOGEGRAMMENI
5044	<E*!>	<U1FC8>	GREEK CAPITAL LETTER EPSILON WITH VARIA
5045	<E*~>	<U1FC9>	GREEK CAPITAL LETTER EPSILON WITH OXIA
5046	<Y*!>	<U1FCA>	GREEK CAPITAL LETTER ETA WITH VARIA
5047	<Y*~>	<U1FCB>	GREEK CAPITAL LETTER ETA WITH OXIA
5048	<Y*J>	<U1FCC>	GREEK CAPITAL LETTER ETA WITH PROSGEGRAMMENI
5049	<, !>	<U1FCD>	GREEK PSILI AND VARIA
5050	<, ~>	<U1FCE>	GREEK PSILI AND OXIA
5051	<?,>	<U1FCF>	GREEK PSILI AND PERISPOMENI
5052	<i*(>	<U1FD0>	GREEK SMALL LETTER IOTA WITH VRACHY
5053	<i*->	<U1FD1>	GREEK SMALL LETTER IOTA WITH MACRON
5054	<i*~!>	<U1FD2>	GREEK SMALL LETTER IOTA WITH DIALYTIKA AND VARIA
5055	<i*~!~>	<U1FD3>	GREEK SMALL LETTER IOTA WITH DIALYTIKA AND OXIA
5056	<i*?>	<U1FD6>	GREEK SMALL LETTER IOTA WITH PERISPOMENI
5057	<i*~?>	<U1FD7>	GREEK SMALL LETTER IOTA WITH DIALYTIKA AND PERISPOMENI
5058	<I*(>	<U1FD8>	GREEK CAPITAL LETTER IOTA WITH VRACHY
5059	<I*->	<U1FD9>	GREEK CAPITAL LETTER IOTA WITH MACRON
5060	<I*~!>	<U1FDA>	GREEK CAPITAL LETTER IOTA WITH VARIA
5061	<I*~!~>	<U1FDB>	GREEK CAPITAL LETTER IOTA WITH OXIA
5062	<!;!>	<U1FDD>	GREEK DASIA AND VARIA
5063	<!~>	<U1FDE>	GREEK DASIA AND OXIA
5064	<?!>	<U1FDF>	GREEK DASIA AND PERISPOMENI
5065	<u*(>	<U1FE0>	GREEK SMALL LETTER UPSILON WITH VRACHY
5066	<u*->	<U1FE1>	GREEK SMALL LETTER UPSILON WITH MACRON
5067	<u*~!>	<U1FE2>	GREEK SMALL LETTER UPSILON WITH DIALYTIKA AND VARIA
5068	<u*~!~>	<U1FE3>	GREEK SMALL LETTER UPSILON WITH DIALYTIKA AND OXIA
5069	<r*~>	<U1FE4>	GREEK SMALL LETTER RHO WITH PSILI
5070	<r*~!>	<U1FE5>	GREEK SMALL LETTER RHO WITH DASIA
5071	<u*?>	<U1FE6>	GREEK SMALL LETTER UPSILON WITH PERISPOMENI
5072	<u*~?>	<U1FE7>	GREEK SMALL LETTER UPSILON WITH DIALYTIKA AND PERISPOMENI
5073	<U*(>	<U1FE8>	GREEK CAPITAL LETTER UPSILON WITH VRACHY
5074	<U*->	<U1FE9>	GREEK CAPITAL LETTER UPSILON WITH MACRON
5075	<U*~!>	<U1FEA>	GREEK CAPITAL LETTER UPSILON WITH VARIA
5076	<U*~!~>	<U1FEB>	GREEK CAPITAL LETTER UPSILON WITH OXIA
5077	<R*~!>	<U1FEC>	GREEK CAPITAL LETTER RHO WITH DASIA
5078	<!;!>	<U1FED>	GREEK DIALYTIKA AND VARIA
5079	<!~>	<U1FEE>	GREEK DIALYTIKA AND OXIA
5080	<!*>	<U1FEF>	GREEK VARIA
5081	<w*!j>	<U1FF2>	GREEK SMALL LETTER OMEGA WITH VARIA AND YPOGEGRAMMENI
5082	<w*j>	<U1FF3>	GREEK SMALL LETTER OMEGA WITH YPOGEGRAMMENI
5083	<w*~j>	<U1FF4>	GREEK SMALL LETTER OMEGA WITH OXIA AND YPOGEGRAMMENI
5084	<w*?>	<U1FF6>	GREEK SMALL LETTER OMEGA WITH PERISPOMENI
5085	<w*~?>	<U1FF7>	GREEK SMALL LETTER OMEGA WITH PERISPOMENI AND YPOGEGRAMMENI
5086	<O*!>	<U1FF8>	GREEK CAPITAL LETTEROMICRON WITH VARIA
5087	<O*~>	<U1FF9>	GREEK CAPITAL LETTEROMICRON WITH OXIA
5088	<W*!>	<U1FFA>	GREEK CAPITAL LETTER OMEGA WITH VARIA
5089	<W*~>	<U1FFB>	GREEK CAPITAL LETTER OMEGA WITH OXIA
5090	<W*J>	<U1FFC>	GREEK CAPITAL LETTER OMEGA WITH PROSGEGRAMMENI
5091	<///>	<U1FFD>	GREEK OXIA
5092	<!/;>	<U1FFE>	GREEK DASIA
5093	<1N>	<U2002>	EN SPACE
5094	<1M>	<U2003>	EM SPACE
5095	<3M>	<U2004>	THREE-PER-EM SPACE
5096	<4M>	<U2005>	FOUR-PER-EM SPACE
5097	<6M>	<U2006>	SIX-PER-EM SPACE
5098	<LR>	<U200E>	LEFT-TO-RIGHT MARK
5099	<RL>	<U200F>	RIGHT-TO-LEFT MARK
5100	<1T>	<U2009>	THIN SPACE
5101	<1H>	<U200A>	HAIR SPACE
5102	<-1>	<U2010>	HYPHEN
5103	<-N>	<U2013>	EN DASH
5104	<-M>	<U2014>	EM DASH
5105	<-3>	<U2015>	HORIZONTAL BAR
5106	<!2>	<U2016>	DOUBLE VERTICAL LINE
5107	<=2>	<U2017>	DOUBLE LOW LINE
5108	<'6>	<U2018>	LEFT SINGLE QUOTATION MARK
5109	<'9>	<U2019>	RIGHT SINGLE QUOTATION MARK
5110	<.9>	<U201A>	SINGLE LOW-9 QUOTATION MARK
5111	<'9>	<U201B>	SINGLE HIGH-REVERSED-9 QUOTATION MARK

112	<"6>	<U201C>	LEFT DOUBLE QUOTATION MARK
113	<"9>	<U201D>	RIGHT DOUBLE QUOTATION MARK
114	<:9>	<U201E>	DOUBLE LOW-9 QUOTATION MARK
115	<9">	<U201F>	DOUBLE HIGH-REVERSED-9 QUOTATION MARK
116	<//->	<U2020>	DAGGER
117	<//=>	<U2021>	DOUBLE DAGGER
118	<sb>	<U2022>	BULLET
119	<3b>	<U2023>	TRIANGULAR BULLET
120	<..>	<U2025>	TWO DOT LEADER
121	<.3>	<U2026>	HORIZONTAL ELLIPSIS
122	<.->	<U2027>	HYPHENATION POINT
123	<linesep>	<U2028>	LINE SEPARATOR
124	<parsep>	<U2029>	PARAGRAPH SEPARATOR
125	<%0>	<U2030>	PER MILLE SIGN
126	<1'>	<U2032>	PRIME
127	<2'>	<U2033>	DOUBLE PRIME
128	<3'>	<U2034>	TRIPLE PRIME
129	<1">	<U2035>	REVERSED PRIME
130	<2">	<U2036>	REVERSED DOUBLE PRIME
131	<3">	<U2037>	REVERSED TRIPLE PRIME
132	<Ca>	<U2038>	CARET
133	<<1>	<U2039>	SINGLE LEFT-POINTING ANGLE QUOTATION MARK
134	</>1>	<U203A>	SINGLE RIGHT-POINTING ANGLE QUOTATION MARK
135	<:X>	<U203B>	REFERENCE MARK
136	<!*2>	<U203C>	DOUBLE EXCLAMATION MARK
137	<' ->	<U203E>	OVERLINE
138	<-b>	<U2043>	HYPHEN BULLET
139	<//f>	<U2044>	FRACTION SLASH
140	<0s>	<U2070>	SUPERSCRIPIT ZERO
141	<4s>	<U2074>	SUPERSCRIPIT FOUR
142	<5s>	<U2075>	SUPERSCRIPIT FIVE
143	<6s>	<U2076>	SUPERSCRIPIT SIX
144	<7s>	<U2077>	SUPERSCRIPIT SEVEN
145	<8s>	<U2078>	SUPERSCRIPIT EIGHT
146	<9s>	<U2079>	SUPERSCRIPIT NINE
147	<+s>	<U207A>	SUPERSCRIPIT PLUS SIGN
148	<-s>	<U207B>	SUPERSCRIPIT MINUS
149	<=s>	<U207C>	SUPERSCRIPIT EQUALS SIGN
150	<(s>	<U207D>	SUPERSCRIPIT LEFT PARENTHESIS
151	<)s>	<U207E>	SUPERSCRIPIT RIGHT PARENTHESIS
152	<nS>	<U207F>	SUPERSCRIPIT LATIN SMALL LETTER N
153	<0s>	<U2080>	SUBSCRIPIT ZERO
154	<1s>	<U2081>	SUBSCRIPIT ONE
155	<2s>	<U2082>	SUBSCRIPIT TWO
156	<3s>	<U2083>	SUBSCRIPIT THREE
157	<4s>	<U2084>	SUBSCRIPIT FOUR
158	<5s>	<U2085>	SUBSCRIPIT FIVE
159	<6s>	<U2086>	SUBSCRIPIT SIX
160	<7s>	<U2087>	SUBSCRIPIT SEVEN
161	<8s>	<U2088>	SUBSCRIPIT EIGHT
162	<9s>	<U2089>	SUBSCRIPIT NINE
163	<+s>	<U208A>	SUBSCRIPIT PLUS SIGN
164	<-s>	<U208B>	SUBSCRIPIT MINUS
165	<=s>	<U208C>	SUBSCRIPIT EQUALS SIGN
166	<(s>	<U208D>	SUBSCRIPIT LEFT PARENTHESIS
167	<)s>	<U208E>	SUBSCRIPIT RIGHT PARENTHESIS
168	<Ff>	<U20A3>	FRENCH FRANC SIGN
169	<Li>	<U20A4>	LIRA SIGN
170	<Pt>	<U20A7>	PESETA SIGN
171	<W>	<U20A9>	WON SIGN
172	<"7>	<U20D1>	COMBINING RIGHT HARPOON ABOVE
173	<oC>	<U2103>	DEGREE CELSIUS
174	<co>	<U2105>	CARE OF
175	<oF>	<U2109>	DEGREE FAHRENHEIT
176	<N0>	<U2116>	NUMERO SIGN
177	<PO>	<U2117>	SOUND RECORDING COPYRIGHT
178	<Rx>	<U211E>	PRESCRIPTION TAKE
179	<SM>	<U2120>	SERVICE MARK
180	<TM>	<U2122>	TRADE MARK SIGN
181	<Om>	<U2126>	OHM SIGN
182	<AO>	<U212B>	ANGSTROM SIGN
183	<Est>	<U212E>	ESTIMATED SYMBOL
184	<13>	<U2153>	VULGAR FRACTION ONE THIRD
185	<23>	<U2154>	VULGAR FRACTION TWO THIRDS
186	<15>	<U2155>	VULGAR FRACTION ONE FIFTH
187	<25>	<U2156>	VULGAR FRACTION TWO FIFTHS
188	<35>	<U2157>	VULGAR FRACTION THREE FIFTHS
189	<45>	<U2158>	VULGAR FRACTION FOUR FIFTHS
190	<16>	<U2159>	VULGAR FRACTION ONE SIXTH
191	<56>	<U215A>	VULGAR FRACTION FIVE SIXTHS
192	<18>	<U215B>	VULGAR FRACTION ONE EIGHTH
193	<38>	<U215C>	VULGAR FRACTION THREE EIGHTHS
194	<58>	<U215D>	VULGAR FRACTION FIVE EIGHTHS
195	<78>	<U215E>	VULGAR FRACTION SEVEN EIGHTHS
196	<1R>	<U2160>	ROMAN NUMERAL ONE
197	<2R>	<U2161>	ROMAN NUMERAL TWO
198	<3R>	<U2162>	ROMAN NUMERAL THREE

5199	<4R>	<U2163>	ROMAN NUMERAL FOUR
5200	<5R>	<U2164>	ROMAN NUMERAL FIVE
5201	<6R>	<U2165>	ROMAN NUMERAL SIX
5202	<7R>	<U2166>	ROMAN NUMERAL SEVEN
5203	<8R>	<U2167>	ROMAN NUMERAL EIGHT
5204	<9R>	<U2168>	ROMAN NUMERAL NINE
5205	<aR>	<U2169>	ROMAN NUMERAL TEN
5206	 	<U216A>	ROMAN NUMERAL ELEVEN
5207	<cR>	<U216B>	ROMAN NUMERAL TWELVE
5208	<50R>	<U216C>	ROMAN NUMERAL FIFTY
5209	<100R>	<U216D>	ROMAN NUMERAL ONE HUNDRED
5210	<500R>	<U216E>	ROMAN NUMERAL FIVE HUNDRED
5211	<1000R>	<U216F>	ROMAN NUMERAL ONE THOUSAND
5212	<1r>	<U2170>	SMALL ROMAN NUMERAL ONE
5213	<2r>	<U2171>	SMALL ROMAN NUMERAL TWO
5214	<3r>	<U2172>	SMALL ROMAN NUMERAL THREE
5215	<4r>	<U2173>	SMALL ROMAN NUMERAL FOUR
5216	<5r>	<U2174>	SMALL ROMAN NUMERAL FIVE
5217	<6r>	<U2175>	SMALL ROMAN NUMERAL SIX
5218	<7r>	<U2176>	SMALL ROMAN NUMERAL SEVEN
5219	<8r>	<U2177>	SMALL ROMAN NUMERAL EIGHT
5220	<9r>	<U2178>	SMALL ROMAN NUMERAL NINE
5221	<ar>	<U2179>	SMALL ROMAN NUMERAL TEN
5222	 	<U217A>	SMALL ROMAN NUMERAL ELEVEN
5223	<cr>	<U217B>	SMALL ROMAN NUMERAL TWELVE
5224	<50r>	<U217C>	SMALL ROMAN NUMERAL FIFTY
5225	<100r>	<U217D>	SMALL ROMAN NUMERAL ONE HUNDRED
5226	<500r>	<U217E>	SMALL ROMAN NUMERAL FIVE HUNDRED
5227	<1000r>	<U217F>	SMALL ROMAN NUMERAL ONE THOUSAND
5228	<1000RCD>	<U2180>	ROMAN NUMERAL ONE THOUSAND C D
5229	<5000R>	<U2181>	ROMAN NUMERAL FIVE THOUSAND
5230	<10000R>	<U2182>	ROMAN NUMERAL TEN THOUSAND
5231	<->	<U2190>	LEFTWARDS ARROW
5232	<-!>	<U2191>	UPWARDS ARROW
5233	<-/>>	<U2192>	RIGHTWARDS ARROW
5234	<-v>	<U2193>	DOWNWARDS ARROW
5235	<</>>	<U2194>	LEFT RIGHT ARROW
5236	<UD>	<U2195>	UP DOWN ARROW
5237	<<!!>	<U2196>	NORTH WEST ARROW
5238	</////>>	<U2197>	NORTH EAST ARROW
5239	<!!/>>	<U2198>	SOUTH EAST ARROW
5240	<</////>	<U2199>	SOUTH WEST ARROW
5241	<UD->	<U21A8>	UP DOWN ARROW WITH BASE
5242	</>V>	<U21C0>	RIGHTWARDS HARPOON WITH BARB UPWARDS
5243	<<=>	<U21D0>	LEFTWARDS DOUBLE ARROW
5244	<=>/>>	<U21D2>	RIGHTWARDS DOUBLE ARROW
5245	<==>	<U21D4>	LEFT RIGHT DOUBLE ARROW
5246	<FA>	<U2200>	FOR ALL
5247	<dP>	<U2202>	PARTIAL DIFFERENTIAL
5248	<TE>	<U2203>	THERE EXISTS
5249	</ / 0>	<U2205>	EMPTY SET
5250	<DE>	<U2206>	INCREMENT
5251	<NB>	<U2207>	NABLA
5252	<(->	<U2208>	ELEMENT OF
5253	<->	<U220B>	CONTAINS AS MEMBER
5254	<FP>	<U220E>	END OF PROOF
5255	<*P>	<U220F>	N-ARY PRODUCT
5256	<+Z>	<U2211>	N-ARY SUMMATION
5257	<-2>	<U2212>	MINUS SIGN
5258	<-+>	<U2213>	MINUS-OR-PLUS SIGN
5259	<.+>	<U2214>	DOT PLUS
5260	<*->	<U2217>	ASTERISK OPERATOR
5261	<Ob>	<U2218>	RING OPERATOR
5262	<Sb>	<U2219>	BULLET OPERATOR
5263	<RT>	<U221A>	SQUARE ROOT
5264	<0(>	<U221D>	PROPORTIONAL TO
5265	<00>	<U221E>	INFINITY
5266	<-L>	<U221F>	RIGHT ANGLE
5267	<-V>	<U2220>	ANGLE
5268	<PP>	<U2225>	PARALLEL TO
5269	<AN>	<U2227>	LOGICAL AND
5270	<OR>	<U2228>	LOGICAL OR
5271	<(U>	<U2229>	INTERSECTION
5272	<)U>	<U222A>	UNION
5273	<In>	<U222B>	INTEGRAL
5274	<DI>	<U222C>	DOUBLE INTEGRAL
5275	<Io>	<U222E>	CONTOUR INTEGRAL
5276	<.:>	<U2234>	THEREFORE
5277	<:.>	<U2235>	BECAUSE
5278	<:R>	<U2236>	RATIO
5279	<::>	<U2237>	PROPORTION
5280	<?1>	<U223C>	TILDE OPERATOR
5281	<CG>	<U223E>	INVERTED LAZY S
5282	<?->	<U2243>	ASYMPTOTICALLY EQUAL TO
5283	<?=>	<U2245>	APPROXIMATELY EQUAL TO
5284	<?2>	<U2248>	ALMOST EQUAL TO
5285	<=?>	<U224C>	ALL EQUAL TO
5286	<HI>	<U2253>	IMAGE OF OR APPROXIMATELY EQUAL TO
5287	<!=>	<U2260>	NOT EQUAL TO

5288	<=3>	<U2261>	IDENTICAL TO
5289	<=<>	<U2264>	LESS-THAN OR EQUAL TO
5290	</>=>	<U2265>	GREATER-THAN OR EQUAL TO
5291	<<*>	<U226A>	MUCH LESS-THAN
5292	<*/>>	<U226B>	MUCH GREATER-THAN
5293	<!<>	<U226E>	NOT LESS-THAN
5294	<!/>>	<U226F>	NOT GREATER-THAN
5295	<(C>	<U2282>	SUBSET OF
5296	<)C>	<U2283>	SUPERSET OF
5297	<( _>	<U2286>	SUBSET OF OR EQUAL TO
5298	<) _>	<U2287>	SUPERSET OF OR EQUAL TO
5299	<0.>	<U2299>	CIRCLED DOT OPERATOR
5300	<02>	<U229A>	CIRCLED RING OPERATOR
5301	<-T>	<U22A5>	UP TACK
5302	<.P>	<U22C5>	DOT OPERATOR
5303	<:3>	<U22EE>	VERTICAL ELLIPSIS
5304	<Eh>	<U2302>	HOUSE
5305	<<7>	<U2308>	LEFT CEILING
5306	</>7>	<U2309>	RIGHT CEILING
5307	<7<>	<U230A>	LEFT FLOOR
5308	<7/>>	<U230B>	RIGHT FLOOR
5309	<NI>	<U2310>	REVERSED NOT SIGN
5310	<(A>	<U2312>	ARC
5311	<TR>	<U2315>	TELEPHONE RECORDER
5312	<88>	<U2318>	PLACE OF INTEREST SIGN
5313	<Iu>	<U2320>	TOP HALF INTEGRAL
5314	<Il>	<U2321>	BOTTOM HALF INTEGRAL
5315	<</>>	<U2329>	LEFT-POINTING ANGLE BRACKET
5316	<///>>	<U232A>	RIGHT-POINTING ANGLE BRACKET
5317	<Vs>	<U2423>	OPEN BOX
5318	<lh>	<U2440>	OCR HOOK
5319	<3h>	<U2441>	OCR CHAIR
5320	<2h>	<U2442>	OCR FORK
5321	<4h>	<U2443>	OCR INVERTED FORK
5322	<1j>	<U2446>	OCR BRANCH BANK IDENTIFICATION
5323	<2j>	<U2447>	OCR AMOUNT OF CHECK
5324	<3j>	<U2448>	OCR DASH
5325	<4j>	<U2449>	OCR CUSTOMER ACCOUNT NUMBER
5326	<1-o>	<U2460>	CIRCLED DIGIT ONE
5327	<2-o>	<U2461>	CIRCLED DIGIT TWO
5328	<3-o>	<U2462>	CIRCLED DIGIT THREE
5329	<4-o>	<U2463>	CIRCLED DIGIT FOUR
5330	<5-o>	<U2464>	CIRCLED DIGIT FIVE
5331	<6-o>	<U2465>	CIRCLED DIGIT SIX
5332	<7-o>	<U2466>	CIRCLED DIGIT SEVEN
5333	<8-o>	<U2467>	CIRCLED DIGIT EIGHT
5334	<9-o>	<U2468>	CIRCLED DIGIT NINE
5335	<10-o>	<U2469>	CIRCLED NUMBER TEN
5336	<11-o>	<U246A>	CIRCLED NUMBER ELEVEN
5337	<12-o>	<U246B>	CIRCLED NUMBER TWELVE
5338	<13-o>	<U246C>	CIRCLED NUMBER THIRTEEN
5339	<14-o>	<U246D>	CIRCLED NUMBER FOURTEEN
5340	<15-o>	<U246E>	CIRCLED NUMBER FIFTEEN
5341	<16-o>	<U246F>	CIRCLED NUMBER SIXTEEN
5342	<17-o>	<U2470>	CIRCLED NUMBER SEVENTEEN
5343	<18-o>	<U2471>	CIRCLED NUMBER EIGHTEEN
5344	<19-o>	<U2472>	CIRCLED NUMBER NINETEEN
5345	<20-o>	<U2473>	CIRCLED NUMBER TWENTY
5346	<(1)>	<U2474>	PARENTHESESIZED DIGIT ONE
5347	<(2)>	<U2475>	PARENTHESESIZED DIGIT TWO
5348	<(3)>	<U2476>	PARENTHESESIZED DIGIT THREE
5349	<(4)>	<U2477>	PARENTHESESIZED DIGIT FOUR
5350	<(5)>	<U2478>	PARENTHESESIZED DIGIT FIVE
5351	<(6)>	<U2479>	PARENTHESESIZED DIGIT SIX
5352	<(7)>	<U247A>	PARENTHESESIZED DIGIT SEVEN
5353	<(8)>	<U247B>	PARENTHESESIZED DIGIT EIGHT
5354	<(9)>	<U247C>	PARENTHESESIZED DIGIT NINE
5355	<(10)>	<U247D>	PARENTHESESIZED NUMBER TEN
5356	<(11)>	<U247E>	PARENTHESESIZED NUMBER ELEVEN
5357	<(12)>	<U247F>	PARENTHESESIZED NUMBER TWELVE
5358	<(13)>	<U2480>	PARENTHESESIZED NUMBER THIRTEEN
5359	<(14)>	<U2481>	PARENTHESESIZED NUMBER FOURTEEN
5360	<(15)>	<U2482>	PARENTHESESIZED NUMBER FIFTEEN
5361	<(16)>	<U2483>	PARENTHESESIZED NUMBER SIXTEEN
5362	<(17)>	<U2484>	PARENTHESESIZED NUMBER SEVENTEEN
5363	<(18)>	<U2485>	PARENTHESESIZED NUMBER EIGHTEEN
5364	<(19)>	<U2486>	PARENTHESESIZED NUMBER NINETEEN
5365	<(20)>	<U2487>	PARENTHESESIZED NUMBER TWENTY
5366	<1.>	<U2488>	DIGIT ONE FULL STOP
5367	<2.>	<U2489>	DIGIT TWO FULL STOP
5368	<3.>	<U248A>	DIGIT THREE FULL STOP
5369	<4.>	<U248B>	DIGIT FOUR FULL STOP
5370	<5.>	<U248C>	DIGIT FIVE FULL STOP
5371	<6.>	<U248D>	DIGIT SIX FULL STOP
5372	<7.>	<U248E>	DIGIT SEVEN FULL STOP
5373	<8.>	<U248F>	DIGIT EIGHT FULL STOP
5374	<9.>	<U2490>	DIGIT NINE FULL STOP

5375	<10.>	<U2491>	NUMBER TEN FULL STOP
5376	<11.>	<U2492>	NUMBER ELEVEN FULL STOP
5377	<12.>	<U2493>	NUMBER TWELVE FULL STOP
5378	<13.>	<U2494>	NUMBER THIRTEEN FULL STOP
5379	<14.>	<U2495>	NUMBER FOURTEEN FULL STOP
5380	<15.>	<U2496>	NUMBER FIFTEEN FULL STOP
5381	<16.>	<U2497>	NUMBER SIXTEEN FULL STOP
5382	<17.>	<U2498>	NUMBER SEVENTEEN FULL STOP
5383	<18.>	<U2499>	NUMBER EIGHTEEN FULL STOP
5384	<19.>	<U249A>	NUMBER NINETEEN FULL STOP
5385	<20.>	<U249B>	NUMBER TWENTY FULL STOP
5386	<(a)>	<U249C>	PARENTHESESIZED LATIN SMALL LETTER A
5387	<(b)>	<U249D>	PARENTHESESIZED LATIN SMALL LETTER B
5388	<(c)>	<U249E>	PARENTHESESIZED LATIN SMALL LETTER C
5389	<(d)>	<U249F>	PARENTHESESIZED LATIN SMALL LETTER D
5390	<(e)>	<U24A0>	PARENTHESESIZED LATIN SMALL LETTER E
5391	<(f)>	<U24A1>	PARENTHESESIZED LATIN SMALL LETTER F
5392	<(g)>	<U24A2>	PARENTHESESIZED LATIN SMALL LETTER G
5393	<(h)>	<U24A3>	PARENTHESESIZED LATIN SMALL LETTER H
5394	<(i)>	<U24A4>	PARENTHESESIZED LATIN SMALL LETTER I
5395	<(j)>	<U24A5>	PARENTHESESIZED LATIN SMALL LETTER J
5396	<(k)>	<U24A6>	PARENTHESESIZED LATIN SMALL LETTER K
5397	<(l)>	<U24A7>	PARENTHESESIZED LATIN SMALL LETTER L
5398	<(m)>	<U24A8>	PARENTHESESIZED LATIN SMALL LETTER M
5399	<(n)>	<U24A9>	PARENTHESESIZED LATIN SMALL LETTER N
5400	<(o)>	<U24AA>	PARENTHESESIZED LATIN SMALL LETTER O
5401	<(p)>	<U24AB>	PARENTHESESIZED LATIN SMALL LETTER P
5402	<(q)>	<U24AC>	PARENTHESESIZED LATIN SMALL LETTER Q
5403	<(r)>	<U24AD>	PARENTHESESIZED LATIN SMALL LETTER R
5404	<(s)>	<U24AE>	PARENTHESESIZED LATIN SMALL LETTER S
5405	<(t)>	<U24AF>	PARENTHESESIZED LATIN SMALL LETTER T
5406	<(u)>	<U24B0>	PARENTHESESIZED LATIN SMALL LETTER U
5407	<(v)>	<U24B1>	PARENTHESESIZED LATIN SMALL LETTER V
5408	<(w)>	<U24B2>	PARENTHESESIZED LATIN SMALL LETTER W
5409	<(x)>	<U24B3>	PARENTHESESIZED LATIN SMALL LETTER X
5410	<(y)>	<U24B4>	PARENTHESESIZED LATIN SMALL LETTER Y
5411	<(z)>	<U24B5>	PARENTHESESIZED LATIN SMALL LETTER Z
5412	<A-o>	<U24B6>	CIRCLED LATIN CAPITAL LETTER A
5413	<B-o>	<U24B7>	CIRCLED LATIN CAPITAL LETTER B
5414	<C-o>	<U24B8>	CIRCLED LATIN CAPITAL LETTER C
5415	<D-o>	<U24B9>	CIRCLED LATIN CAPITAL LETTER D
5416	<E-o>	<U24BA>	CIRCLED LATIN CAPITAL LETTER E
5417	<F-o>	<U24BB>	CIRCLED LATIN CAPITAL LETTER F
5418	<G-o>	<U24BC>	CIRCLED LATIN CAPITAL LETTER G
5419	<H-o>	<U24BD>	CIRCLED LATIN CAPITAL LETTER H
5420	<I-o>	<U24BE>	CIRCLED LATIN CAPITAL LETTER I
5421	<J-o>	<U24BF>	CIRCLED LATIN CAPITAL LETTER J
5422	<K-o>	<U24C0>	CIRCLED LATIN CAPITAL LETTER K
5423	<L-o>	<U24C1>	CIRCLED LATIN CAPITAL LETTER L
5424	<M-o>	<U24C2>	CIRCLED LATIN CAPITAL LETTER M
5425	<N-o>	<U24C3>	CIRCLED LATIN CAPITAL LETTER N
5426	<O-o>	<U24C4>	CIRCLED LATIN CAPITAL LETTER O
5427	<P-o>	<U24C5>	CIRCLED LATIN CAPITAL LETTER P
5428	<Q-o>	<U24C6>	CIRCLED LATIN CAPITAL LETTER Q
5429	<R-o>	<U24C7>	CIRCLED LATIN CAPITAL LETTER R
5430	<S-o>	<U24C8>	CIRCLED LATIN CAPITAL LETTER S
5431	<T-o>	<U24C9>	CIRCLED LATIN CAPITAL LETTER T
5432	<U-o>	<U24CA>	CIRCLED LATIN CAPITAL LETTER U
5433	<V-o>	<U24CB>	CIRCLED LATIN CAPITAL LETTER V
5434	<W-o>	<U24CC>	CIRCLED LATIN CAPITAL LETTER W
5435	<X-o>	<U24CD>	CIRCLED LATIN CAPITAL LETTER X
5436	<Y-o>	<U24CE>	CIRCLED LATIN CAPITAL LETTER Y
5437	<Z-o>	<U24CF>	CIRCLED LATIN CAPITAL LETTER Z
5438	<a-o>	<U24D0>	CIRCLED LATIN SMALL LETTER A
5439	<b-o>	<U24D1>	CIRCLED LATIN SMALL LETTER B
5440	<c-o>	<U24D2>	CIRCLED LATIN SMALL LETTER C
5441	<d-o>	<U24D3>	CIRCLED LATIN SMALL LETTER D
5442	<e-o>	<U24D4>	CIRCLED LATIN SMALL LETTER E
5443	<f-o>	<U24D5>	CIRCLED LATIN SMALL LETTER F
5444	<g-o>	<U24D6>	CIRCLED LATIN SMALL LETTER G
5445	<h-o>	<U24D7>	CIRCLED LATIN SMALL LETTER H
5446	<i-o>	<U24D8>	CIRCLED LATIN SMALL LETTER I
5447	<j-o>	<U24D9>	CIRCLED LATIN SMALL LETTER J
5448	<k-o>	<U24DA>	CIRCLED LATIN SMALL LETTER K
5449	<l-o>	<U24DB>	CIRCLED LATIN SMALL LETTER L
5450	<m-o>	<U24DC>	CIRCLED LATIN SMALL LETTER M
5451	<n-o>	<U24DD>	CIRCLED LATIN SMALL LETTER N
5452	<o-o>	<U24DE>	CIRCLED LATIN SMALL LETTER O
5453	<p-o>	<U24DF>	CIRCLED LATIN SMALL LETTER P
5454	<q-o>	<U24E0>	CIRCLED LATIN SMALL LETTER Q
5455	<r-o>	<U24E1>	CIRCLED LATIN SMALL LETTER R
5456	<s-o>	<U24E2>	CIRCLED LATIN SMALL LETTER S
5457	<t-o>	<U24E3>	CIRCLED LATIN SMALL LETTER T
5458	<u-o>	<U24E4>	CIRCLED LATIN SMALL LETTER U
5459	<v-o>	<U24E5>	CIRCLED LATIN SMALL LETTER V
5460	<w-o>	<U24E6>	CIRCLED LATIN SMALL LETTER W
5461	<x-o>	<U24E7>	CIRCLED LATIN SMALL LETTER X
5462	<y-o>	<U24E8>	CIRCLED LATIN SMALL LETTER Y
5463	<z-o>	<U24E9>	CIRCLED LATIN SMALL LETTER Z

5464	<0-o>	<U24EA>	CIRCLED DIGIT ZERO
5465	<hh>	<U2500>	BOX DRAWINGS LIGHT HORIZONTAL
5466	<HH->	<U2501>	BOX DRAWINGS HEAVY HORIZONTAL
5467	<vv>	<U2502>	BOX DRAWINGS LIGHT VERTICAL
5468	<VV->	<U2503>	BOX DRAWINGS HEAVY VERTICAL
5469	<3->	<U2504>	BOX DRAWINGS LIGHT TRIPLE DASH HORIZONTAL
5470	<3_>	<U2505>	BOX DRAWINGS HEAVY TRIPLE DASH HORIZONTAL
5471	<3!>	<U2506>	BOX DRAWINGS LIGHT TRIPLE DASH VERTICAL
5472	<3//>	<U2507>	BOX DRAWINGS HEAVY TRIPLE DASH VERTICAL
5473	<4->	<U2508>	BOX DRAWINGS LIGHT QUADRUPLE DASH HORIZONTAL
5474	<4_>	<U2509>	BOX DRAWINGS HEAVY QUADRUPLE DASH HORIZONTAL
5475	<4!>	<U250A>	BOX DRAWINGS LIGHT QUADRUPLE DASH VERTICAL
5476	<4//>	<U250B>	BOX DRAWINGS HEAVY QUADRUPLE DASH VERTICAL
5477	<dr>	<U250C>	BOX DRAWINGS LIGHT DOWN AND RIGHT
5478	<dR->	<U250D>	BOX DRAWINGS DOWN LIGHT AND RIGHT HEAVY
5479	<Dr->	<U250E>	BOX DRAWINGS DOWN HEAVY AND RIGHT LIGHT
5480	<DR->	<U250F>	BOX DRAWINGS HEAVY DOWN AND RIGHT
5481	<dl>	<U2510>	BOX DRAWINGS LIGHT DOWN AND LEFT
5482	<dL->	<U2511>	BOX DRAWINGS DOWN LIGHT AND LEFT HEAVY
5483	<Dl->	<U2512>	BOX DRAWINGS DOWN HEAVY AND LEFT LIGHT
5484	<LD->	<U2513>	BOX DRAWINGS HEAVY DOWN AND LEFT
5485	<ur>	<U2514>	BOX DRAWINGS LIGHT UP AND RIGHT
5486	<uR->	<U2515>	BOX DRAWINGS UP LIGHT AND RIGHT HEAVY
5487	<Ur->	<U2516>	BOX DRAWINGS UP HEAVY AND RIGHT LIGHT
5488	<UR->	<U2517>	BOX DRAWINGS HEAVY UP AND RIGHT
5489	<ul>	<U2518>	BOX DRAWINGS LIGHT UP AND LEFT
5490	<uL->	<U2519>	BOX DRAWINGS UP LIGHT AND LEFT HEAVY
5491	<Ul->	<U251A>	BOX DRAWINGS UP HEAVY AND LEFT LIGHT
5492	<UL->	<U251B>	BOX DRAWINGS HEAVY UP AND LEFT
5493	<vr>	<U251C>	BOX DRAWINGS LIGHT VERTICAL AND RIGHT
5494	<vR->	<U251D>	BOX DRAWINGS VERTICAL LIGHT AND RIGHT HEAVY
5495	<Udr>	<U251E>	BOX DRAWINGS UP HEAVY AND RIGHT DOWN LIGHT
5496	<uDr>	<U251F>	BOX DRAWINGS DOWN HEAVY AND RIGHT UP LIGHT
5497	<Vr->	<U2520>	BOX DRAWINGS VERTICAL HEAVY AND RIGHT LIGHT
5498	<UdR>	<U2521>	BOX DRAWINGS DOWN LIGHT AND RIGHT UP HEAVY
5499	<uDR>	<U2522>	BOX DRAWINGS UP LIGHT AND RIGHT DOWN HEAVY
5500	<VR->	<U2523>	BOX DRAWINGS HEAVY VERTICAL AND RIGHT
5501	<vl>	<U2524>	BOX DRAWINGS LIGHT VERTICAL AND LEFT
5502	<vL->	<U2525>	BOX DRAWINGS VERTICAL LIGHT AND LEFT HEAVY
5503	<Udl>	<U2526>	BOX DRAWINGS UP HEAVY AND LEFT DOWN LIGHT
5504	<uDl>	<U2527>	BOX DRAWINGS DOWN HEAVY AND LEFT UP LIGHT
5505	<Vl->	<U2528>	BOX DRAWINGS VERTICAL HEAVY AND LEFT LIGHT
5506	<UdL>	<U2529>	BOX DRAWINGS DOWN LIGHT AND LEFT UP HEAVY
5507	<uDl>	<U252A>	BOX DRAWINGS UP LIGHT AND LEFT DOWN HEAVY
5508	<VL->	<U252B>	BOX DRAWINGS HEAVY VERTICAL AND LEFT
5509	<dh>	<U252C>	BOX DRAWINGS LIGHT DOWN AND HORIZONTAL
5510	<dLr>	<U252D>	BOX DRAWINGS LEFT HEAVY AND RIGHT DOWN LIGHT
5511	<dLr>	<U252E>	BOX DRAWINGS RIGHT HEAVY AND LEFT DOWN LIGHT
5512	<dH->	<U252F>	BOX DRAWINGS DOWN LIGHT AND HORIZONTAL HEAVY
5513	<Dh->	<U2530>	BOX DRAWINGS DOWN HEAVY AND HORIZONTAL LIGHT
5514	<DLr>	<U2531>	BOX DRAWINGS RIGHT LIGHT AND LEFT DOWN HEAVY
5515	<DLr>	<U2532>	BOX DRAWINGS LEFT LIGHT AND RIGHT DOWN HEAVY
5516	<DH->	<U2533>	BOX DRAWINGS HEAVY DOWN AND HORIZONTAL
5517	<uh>	<U2534>	BOX DRAWINGS LIGHT UP AND HORIZONTAL
5518	<uLr>	<U2535>	BOX DRAWINGS LEFT HEAVY AND RIGHT UP LIGHT
5519	<uLr>	<U2536>	BOX DRAWINGS RIGHT HEAVY AND LEFT UP LIGHT
5520	<uH->	<U2537>	BOX DRAWINGS UP LIGHT AND HORIZONTAL HEAVY
5521	<Uh->	<U2538>	BOX DRAWINGS UP HEAVY AND HORIZONTAL LIGHT
5522	<ULr>	<U2539>	BOX DRAWINGS RIGHT LIGHT AND LEFT UP HEAVY
5523	<ULr>	<U253A>	BOX DRAWINGS LEFT LIGHT AND RIGHT UP HEAVY
5524	<UH->	<U253B>	BOX DRAWINGS HEAVY UP AND HORIZONTAL
5525	<vh>	<U253C>	BOX DRAWINGS LIGHT VERTICAL AND HORIZONTAL
5526	<vLr>	<U253D>	BOX DRAWINGS LEFT HEAVY AND RIGHT VERTICAL LIGHT
5527	<vLr>	<U253E>	BOX DRAWINGS RIGHT HEAVY AND LEFT VERTICAL LIGHT
5528	<vH->	<U253F>	BOX DRAWINGS VERTICAL LIGHT AND HORIZONTAL HEAVY
5529	<Udh>	<U2540>	BOX DRAWINGS UP HEAVY AND DOWN HORIZONTAL LIGHT
5530	<uDh>	<U2541>	BOX DRAWINGS DOWN HEAVY AND UP HORIZONTAL LIGHT
5531	<Vh->	<U2542>	BOX DRAWINGS VERTICAL HEAVY AND HORIZONTAL LIGHT
5532	<UdLr>	<U2543>	BOX DRAWINGS LEFT UP HEAVY AND RIGHT DOWN LIGHT
5533	<UdLr>	<U2544>	BOX DRAWINGS RIGHT UP HEAVY AND LEFT DOWN LIGHT
5534	<uDlR>	<U2545>	BOX DRAWINGS LEFT DOWN HEAVY AND RIGHT UP LIGHT
5535	<uDlR>	<U2546>	BOX DRAWINGS RIGHT DOWN HEAVY AND LEFT UP LIGHT
5536	<UdH>	<U2547>	BOX DRAWINGS DOWN LIGHT AND UP HORIZONTAL HEAVY
5537	<uDh>	<U2548>	BOX DRAWINGS UP LIGHT AND DOWN HORIZONTAL HEAVY
5538	<VLr>	<U2549>	BOX DRAWINGS RIGHT LIGHT AND LEFT VERTICAL HEAVY
5539	<VlR>	<U254A>	BOX DRAWINGS LEFT LIGHT AND RIGHT VERTICAL HEAVY
5540	<VH->	<U254B>	BOX DRAWINGS HEAVY VERTICAL AND HORIZONTAL
5541	<HH>	<U2550>	BOX DRAWINGS DOUBLE HORIZONTAL
5542	<VV>	<U2551>	BOX DRAWINGS DOUBLE VERTICAL
5543	<dR>	<U2552>	BOX DRAWINGS DOWN SINGLE AND RIGHT DOUBLE
5544	<Dr>	<U2553>	BOX DRAWINGS DOWN DOUBLE AND RIGHT SINGLE
5545	<DR>	<U2554>	BOX DRAWINGS DOUBLE DOWN AND RIGHT
5546	<dL>	<U2555>	BOX DRAWINGS DOWN SINGLE AND LEFT DOUBLE
5547	<Dl>	<U2556>	BOX DRAWINGS DOWN DOUBLE AND LEFT SINGLE
5548	<LD>	<U2557>	BOX DRAWINGS DOUBLE DOWN AND LEFT
5549	<uR>	<U2558>	BOX DRAWINGS UP SINGLE AND RIGHT DOUBLE
5550	<Ur>	<U2559>	BOX DRAWINGS UP DOUBLE AND RIGHT SINGLE



5551	<UR>	<U255A>	BOX DRAWINGS DOUBLE UP AND RIGHT
5552	<uL>	<U255B>	BOX DRAWINGS UP SINGLE AND LEFT DOUBLE
5553	<U1>	<U255C>	BOX DRAWINGS UP DOUBLE AND LEFT SINGLE
5554	<UL>	<U255D>	BOX DRAWINGS DOUBLE UP AND LEFT
5555	<vR>	<U255E>	BOX DRAWINGS VERTICAL SINGLE AND RIGHT DOUBLE
5556	<Vr>	<U255F>	BOX DRAWINGS VERTICAL DOUBLE AND RIGHT SINGLE
5557	<VR>	<U2560>	BOX DRAWINGS DOUBLE VERTICAL AND RIGHT
5558	<vL>	<U2561>	BOX DRAWINGS VERTICAL SINGLE AND LEFT DOUBLE
5559	<Vl>	<U2562>	BOX DRAWINGS VERTICAL DOUBLE AND LEFT SINGLE
5560	<VL>	<U2563>	BOX DRAWINGS DOUBLE VERTICAL AND LEFT
5561	<dH>	<U2564>	BOX DRAWINGS DOWN SINGLE AND HORIZONTAL DOUBLE
5562	<Dh>	<U2565>	BOX DRAWINGS DOWN DOUBLE AND HORIZONTAL SINGLE
5563	<DH>	<U2566>	BOX DRAWINGS DOUBLE DOWN AND HORIZONTAL
5564	<uH>	<U2567>	BOX DRAWINGS UP SINGLE AND HORIZONTAL DOUBLE
5565	<Uh>	<U2568>	BOX DRAWINGS UP DOUBLE AND HORIZONTAL SINGLE
5566	<UH>	<U2569>	BOX DRAWINGS DOUBLE UP AND HORIZONTAL
5567	<vH>	<U256A>	BOX DRAWINGS VERTICAL SINGLE AND HORIZONTAL DOUBLE
5568	<Vh>	<U256B>	BOX DRAWINGS VERTICAL DOUBLE AND HORIZONTAL SINGLE
5569	<VH>	<U256C>	BOX DRAWINGS DOUBLE VERTICAL AND HORIZONTAL
5570	<FD>	<U2571>	BOX DRAWINGS LIGHT DIAGONAL UPPER RIGHT TO LOWER LEFT
5571	<BD>	<U2572>	BOX DRAWINGS LIGHT DIAGONAL UPPER LEFT TO LOWER RIGHT
5572	<TB>	<U2580>	UPPER HALF BLOCK
5573	<LB>	<U2584>	LOWER HALF BLOCK
5574	<FB>	<U2588>	FULL BLOCK
5575	<LB>	<U258C>	LEFT HALF BLOCK
5576	<RB>	<U2590>	RIGHT HALF BLOCK
5577	<.S>	<U2591>	LIGHT SHADE
5578	<.:S>	<U2592>	MEDIUM SHADE
5579	<?S>	<U2593>	DARK SHADE
5580	<fS>	<U25A0>	BLACK SQUARE
5581	<oS>	<U25A1>	WHITE SQUARE
5582	<ro>	<U25A2>	WHITE SQUARE WITH ROUNDED CORNERS
5583	<Rr>	<U25A3>	WHITE SQUARE CONTAINING BLACK SMALL SQUARE
5584	<RF>	<U25A4>	SQUARE WITH HORIZONTAL FILL
5585	<RY>	<U25A5>	SQUARE WITH VERTICAL FILL
5586	<RH>	<U25A6>	SQUARE WITH ORTHOGONAL CROSSHATCH FILL
5587	<RZ>	<U25A7>	SQUARE WITH UPPER LEFT TO LOWER RIGHT FILL
5588	<RK>	<U25A8>	SQUARE WITH UPPER RIGHT TO LOWER LEFT FILL
5589	<RX>	<U25A9>	SQUARE WITH DIAGONAL CROSSHATCH FILL
5590	<sB>	<U25AA>	BLACK SMALL SQUARE
5591	<SR>	<U25AC>	BLACK RECTANGLE
5592	<Or>	<U25AD>	WHITE RECTANGLE
5593	<UT>	<U25B2>	BLACK UP-POINTING TRIANGLE
5594	<uT>	<U25B3>	WHITE UP-POINTING TRIANGLE
5595	<Tr>	<U25B7>	WHITE RIGHT-POINTING TRIANGLE
5596	<PR>	<U25BA>	BLACK RIGHT-POINTING POINTER
5597	<Dt>	<U25BC>	BLACK DOWN-POINTING TRIANGLE
5598	<dT>	<U25BD>	WHITE DOWN-POINTING TRIANGLE
5599	<Tl>	<U25C1>	WHITE LEFT-POINTING TRIANGLE
5600	<PL>	<U25C4>	BLACK LEFT-POINTING POINTER
5601	<Db>	<U25C6>	BLACK DIAMOND
5602	<Dw>	<U25C7>	WHITE DIAMOND
5603	<LZ>	<U25CA>	LOZENGE
5604	<Om>	<U25CB>	WHITE CIRCLE
5605	<Oo>	<U25CE>	BULLSEYE
5606	<OM>	<U25CF>	BLACK CIRCLE
5607	<OL>	<U25D0>	CIRCLE WITH LEFT HALF BLACK
5608	<OR>	<U25D1>	CIRCLE WITH RIGHT HALF BLACK
5609	<Sn>	<U25D8>	INVERSE BULLET
5610	<Ic>	<U25D9>	INVERSE WHITE CIRCLE
5611	<Fd>	<U25E2>	BLACK LOWER RIGHT TRIANGLE
5612	<Bd>	<U25E3>	BLACK LOWER LEFT TRIANGLE
5613	<Ci>	<U25EF>	LARGE CIRCLE
5614	<*2>	<U2605>	BLACK STAR
5615	<*1>	<U2606>	WHITE STAR
5616	<TEL>	<U260E>	BLACK TELEPHONE
5617	<tel>	<U260F>	WHITE TELEPHONE
5618	<<H>	<U261C>	WHITE LEFT POINTING INDEX
5619	</>H>	<U261E>	WHITE RIGHT POINTING INDEX
5620	<Ou>	<U263A>	WHITE SMILING FACE
5621	<OU>	<U263B>	BLACK SMILING FACE
5622	<SU>	<U263C>	WHITE SUN WITH RAYS
5623	<Fm>	<U2640>	FEMALE SIGN
5624	<Ml>	<U2642>	MALE SIGN
5625	<cS>	<U2660>	BLACK SPADE SUIT
5626	<ch>	<U2661>	WHITE HEART SUIT
5627	<cd>	<U2662>	WHITE DIAMOND SUIT
5628	<cC>	<U2663>	BLACK CLUB SUIT
5629	<cS->	<U2664>	WHITE SPADE SUIT
5630	<cH->	<U2665>	BLACK HEART SUIT
5631	<cD->	<U2666>	BLACK DIAMOND SUIT
5632	<cC->	<U2667>	WHITE CLUB SUIT
5633	<Md>	<U2669>	QUARTER NOTE
5634	<M8>	<U266A>	EIGHTH NOTE
5635	<M2>	<U266B>	BEAMED EIGHTH NOTES
5636	<M16>	<U266C>	BEAMED SIXTEENTH NOTES
5637	<Mb>	<U266D>	MUSIC FLAT SIGN
5638	<Mx>	<U266E>	MUSIC NATURAL SIGN
5639	<MX>	<U266F>	MUSIC SHARP SIGN

5640	<OK>	<U2713>	CHECK MARK
5641	<XX>	<U2717>	BALLOT X
5642	<-X>	<U2720>	MALTESE CROSS
5643	<IS>	<U3000>	IDEOGRAPHIC SPACE
5644	< , _>	<U3001>	IDEOGRAPHIC COMMA
5645	< . _>	<U3002>	IDEOGRAPHIC FULL STOP
5646	<+>	<U3003>	DITTO MARK
5647	<JIS>	<U3004>	JAPANESE INDUSTRIAL STANDARD SYMBOL
5648	<* _>	<U3005>	IDEOGRAPHIC ITERATION MARK
5649	< ; _>	<U3006>	IDEOGRAPHIC CLOSING MARK
5650	<0 _>	<U3007>	IDEOGRAPHIC NUMBER ZERO
5651	<<+>	<U300A>	LEFT DOUBLE ANGLE BRACKET
5652	</>+>	<U300B>	RIGHT DOUBLE ANGLE BRACKET
5653	<<'>	<U300C>	LEFT CORNER BRACKET
5654	</>'>	<U300D>	RIGHT CORNER BRACKET
5655	<<">	<U300E>	LEFT WHITE CORNER BRACKET
5656	</>">	<U300F>	RIGHT WHITE CORNER BRACKET
5657	<(">	<U3010>	LEFT BLACK LENTICULAR BRACKET
5658	<)">	<U3011>	RIGHT BLACK LENTICULAR BRACKET
5659	<=T>	<U3012>	POSTAL MARK
5660	<=_>	<U3013>	GETA MARK
5661	<('>	<U3014>	LEFT TORTOISE SHELL BRACKET
5662	<)'>	<U3015>	RIGHT TORTOISE SHELL BRACKET
5663	<(I>	<U3016>	LEFT WHITE LENTICULAR BRACKET
5664	<)I>	<U3017>	RIGHT WHITE LENTICULAR BRACKET
5665	<-?>	<U301C>	WAVE DASH
5666	<=T:)>	<U3020>	POSTAL MARK FACE
5667	<A5>	<U3041>	HIRAGANA LETTER SMALL A
5668	<a5>	<U3042>	HIRAGANA LETTER A
5669	<I5>	<U3043>	HIRAGANA LETTER SMALL I
5670	<i5>	<U3044>	HIRAGANA LETTER I
5671	<U5>	<U3045>	HIRAGANA LETTER SMALL U
5672	<u5>	<U3046>	HIRAGANA LETTER U
5673	<E5>	<U3047>	HIRAGANA LETTER SMALL E
5674	<e5>	<U3048>	HIRAGANA LETTER E
5675	<O5>	<U3049>	HIRAGANA LETTER SMALL O
5676	<o5>	<U304A>	HIRAGANA LETTER O
5677	<ka>	<U304B>	HIRAGANA LETTER KA
5678	<ga>	<U304C>	HIRAGANA LETTER GA
5679	<ki>	<U304D>	HIRAGANA LETTER KI
5680	<gi>	<U304E>	HIRAGANA LETTER GI
5681	<ku>	<U304F>	HIRAGANA LETTER KU
5682	<gu>	<U3050>	HIRAGANA LETTER GU
5683	<ke>	<U3051>	HIRAGANA LETTER KE
5684	<ge>	<U3052>	HIRAGANA LETTER GE
5685	<ko>	<U3053>	HIRAGANA LETTER KO
5686	<go>	<U3054>	HIRAGANA LETTER GO
5687	<sa>	<U3055>	HIRAGANA LETTER SA
5688	<za>	<U3056>	HIRAGANA LETTER ZA
5689	<si>	<U3057>	HIRAGANA LETTER SI
5690	<zi>	<U3058>	HIRAGANA LETTER ZI
5691	<su>	<U3059>	HIRAGANA LETTER SU
5692	<zu>	<U305A>	HIRAGANA LETTER ZU
5693	<se>	<U305B>	HIRAGANA LETTER SE
5694	<ze>	<U305C>	HIRAGANA LETTER ZE
5695	<so>	<U305D>	HIRAGANA LETTER SO
5696	<zo>	<U305E>	HIRAGANA LETTER ZO
5697	<ta>	<U305F>	HIRAGANA LETTER TA
5698	<da>	<U3060>	HIRAGANA LETTER DA
5699	<ti>	<U3061>	HIRAGANA LETTER TI
5700	<di>	<U3062>	HIRAGANA LETTER DI
5701	<tu>	<U3063>	HIRAGANA LETTER SMALL TU
5702	<tu>	<U3064>	HIRAGANA LETTER TU
5703	<du>	<U3065>	HIRAGANA LETTER DU
5704	<te>	<U3066>	HIRAGANA LETTER TE
5705	<de>	<U3067>	HIRAGANA LETTER DE
5706	<to>	<U3068>	HIRAGANA LETTER TO
5707	<do>	<U3069>	HIRAGANA LETTER DO
5708	<na>	<U306A>	HIRAGANA LETTER NA
5709	<ni>	<U306B>	HIRAGANA LETTER NI
5710	<nu>	<U306C>	HIRAGANA LETTER NU
5711	<ne>	<U306D>	HIRAGANA LETTER NE
5712	<no>	<U306E>	HIRAGANA LETTER NO
5713	<ha>	<U306F>	HIRAGANA LETTER HA
5714	<ba>	<U3070>	HIRAGANA LETTER BA
5715	<pa>	<U3071>	HIRAGANA LETTER PA
5716	<hi>	<U3072>	HIRAGANA LETTER HI
5717	<bi>	<U3073>	HIRAGANA LETTER BI
5718	<pi>	<U3074>	HIRAGANA LETTER PI
5719	<hu>	<U3075>	HIRAGANA LETTER HU
5720	<bu>	<U3076>	HIRAGANA LETTER BU
5721	<pu>	<U3077>	HIRAGANA LETTER PU
5722	<he>	<U3078>	HIRAGANA LETTER HE
5723	<be>	<U3079>	HIRAGANA LETTER BE
5724	<pe>	<U307A>	HIRAGANA LETTER PE
5725	<ho>	<U307B>	HIRAGANA LETTER HO
5726	<bo>	<U307C>	HIRAGANA LETTER BO

5727	<po>	<U307D>	HIRAGANA LETTER PO
5728	<ma>	<U307E>	HIRAGANA LETTER MA
5729	<mi>	<U307F>	HIRAGANA LETTER MI
5730	<mu>	<U3080>	HIRAGANA LETTER MU
5731	<me>	<U3081>	HIRAGANA LETTER ME
5732	<mo>	<U3082>	HIRAGANA LETTER MO
5733	<yA>	<U3083>	HIRAGANA LETTER SMALL YA
5734	<ya>	<U3084>	HIRAGANA LETTER YA
5735	<yU>	<U3085>	HIRAGANA LETTER SMALL YU
5736	<yu>	<U3086>	HIRAGANA LETTER YU
5737	<yO>	<U3087>	HIRAGANA LETTER SMALL YO
5738	<yo>	<U3088>	HIRAGANA LETTER YO
5739	<ra>	<U3089>	HIRAGANA LETTER RA
5740	<ri>	<U308A>	HIRAGANA LETTER RI
5741	<ru>	<U308B>	HIRAGANA LETTER RU
5742	<re>	<U308C>	HIRAGANA LETTER RE
5743	<ro>	<U308D>	HIRAGANA LETTER RO
5744	<wA>	<U308E>	HIRAGANA LETTER SMALL WA
5745	<wa>	<U308F>	HIRAGANA LETTER WA
5746	<wi>	<U3090>	HIRAGANA LETTER WI
5747	<we>	<U3091>	HIRAGANA LETTER WE
5748	<wo>	<U3092>	HIRAGANA LETTER WO
5749	<n5>	<U3093>	HIRAGANA LETTER N
5750	<vu>	<U3094>	HIRAGANA LETTER VU
5751	<"5>	<U309B>	KATAKANA-HIRAGANA VOICED SOUND MARK
5752	<05>	<U309C>	KATAKANA-HIRAGANA SEMI-VOICED SOUND MARK
5753	<*5>	<U309D>	HIRAGANA ITERATION MARK
5754	<+5>	<U309E>	HIRAGANA VOICED ITERATION MARK
5755	<a6>	<U30A1>	KATAKANA LETTER SMALL A
5756	<A6>	<U30A2>	KATAKANA LETTER A
5757	<i6>	<U30A3>	KATAKANA LETTER SMALL I
5758	<I6>	<U30A4>	KATAKANA LETTER I
5759	<u6>	<U30A5>	KATAKANA LETTER SMALL U
5760	<U6>	<U30A6>	KATAKANA LETTER U
5761	<e6>	<U30A7>	KATAKANA LETTER SMALL E
5762	<E6>	<U30A8>	KATAKANA LETTER E
5763	<o6>	<U30A9>	KATAKANA LETTER SMALL O
5764	<O6>	<U30AA>	KATAKANA LETTER O
5765	<Ka>	<U30AB>	KATAKANA LETTER KA
5766	<Ga>	<U30AC>	KATAKANA LETTER GA
5767	<Ki>	<U30AD>	KATAKANA LETTER KI
5768	<Gi>	<U30AE>	KATAKANA LETTER GI
5769	<Ku>	<U30AF>	KATAKANA LETTER KU
5770	<Gu>	<U30B0>	KATAKANA LETTER GU
5771	<Ke>	<U30B1>	KATAKANA LETTER KE
5772	<Ge>	<U30B2>	KATAKANA LETTER GE
5773	<Ko>	<U30B3>	KATAKANA LETTER KO
5774	<Go>	<U30B4>	KATAKANA LETTER GO
5775	<Sa>	<U30B5>	KATAKANA LETTER SA
5776	<Za>	<U30B6>	KATAKANA LETTER ZA
5777	<Si>	<U30B7>	KATAKANA LETTER SI
5778	<Zi>	<U30B8>	KATAKANA LETTER ZI
5779	<Su>	<U30B9>	KATAKANA LETTER SU
5780	<Zu>	<U30BA>	KATAKANA LETTER ZU
5781	<Se>	<U30BB>	KATAKANA LETTER SE
5782	<Ze>	<U30BC>	KATAKANA LETTER ZE
5783	<So>	<U30BD>	KATAKANA LETTER SO
5784	<Zo>	<U30BE>	KATAKANA LETTER ZO
5785	<Ta>	<U30BF>	KATAKANA LETTER TA
5786	<Da>	<U30C0>	KATAKANA LETTER DA
5787	<Ti>	<U30C1>	KATAKANA LETTER TI
5788	<Di>	<U30C2>	KATAKANA LETTER DI
5789	<Tu>	<U30C3>	KATAKANA LETTER SMALL TU
5790	<Tu>	<U30C4>	KATAKANA LETTER TU
5791	<Du>	<U30C5>	KATAKANA LETTER DU
5792	<Te>	<U30C6>	KATAKANA LETTER TE
5793	<De>	<U30C7>	KATAKANA LETTER DE
5794	<To>	<U30C8>	KATAKANA LETTER TO
5795	<Do>	<U30C9>	KATAKANA LETTER DO
5796	<Na>	<U30CA>	KATAKANA LETTER NA
5797	<Ni>	<U30CB>	KATAKANA LETTER NI
5798	<Nu>	<U30CC>	KATAKANA LETTER NU
5799	<Ne>	<U30CD>	KATAKANA LETTER NE
5800	<No>	<U30CE>	KATAKANA LETTER NO
5801	<Ha>	<U30CF>	KATAKANA LETTER HA
5802	<Ba>	<U30D0>	KATAKANA LETTER BA
5803	<Pa>	<U30D1>	KATAKANA LETTER PA
5804	<Hi>	<U30D2>	KATAKANA LETTER HI
5805	<Bi>	<U30D3>	KATAKANA LETTER BI
5806	<Pi>	<U30D4>	KATAKANA LETTER PI
5807	<Hu>	<U30D5>	KATAKANA LETTER HU
5808	<Bu>	<U30D6>	KATAKANA LETTER BU
5809	<Pu>	<U30D7>	KATAKANA LETTER PU
5810	<He>	<U30D8>	KATAKANA LETTER HE
5811	<Be>	<U30D9>	KATAKANA LETTER BE
5812	<Pe>	<U30DA>	KATAKANA LETTER PE
5813	<Ho>	<U30DB>	KATAKANA LETTER HO
5814	<Bo>	<U30DC>	KATAKANA LETTER BO
5815	<Po>	<U30DD>	KATAKANA LETTER PO

5816	<Ma>	<U30DE>	KATAKANA LETTER MA
5817	<Mi>	<U30DF>	KATAKANA LETTER MI
5818	<Mu>	<U30E0>	KATAKANA LETTER MU
5819	<Me>	<U30E1>	KATAKANA LETTER ME
5820	<Mo>	<U30E2>	KATAKANA LETTER MO
5821	<YA>	<U30E3>	KATAKANA LETTER SMALL YA
5822	<Ya>	<U30E4>	KATAKANA LETTER YA
5823	<YU>	<U30E5>	KATAKANA LETTER SMALL YU
5824	<Yu>	<U30E6>	KATAKANA LETTER YU
5825	<YO>	<U30E7>	KATAKANA LETTER SMALL YO
5826	<Yo>	<U30E8>	KATAKANA LETTER YO
5827	<Ra>	<U30E9>	KATAKANA LETTER RA
5828	<Ri>	<U30EA>	KATAKANA LETTER RI
5829	<Ru>	<U30EB>	KATAKANA LETTER RU
5830	<Re>	<U30EC>	KATAKANA LETTER RE
5831	<Ro>	<U30ED>	KATAKANA LETTER RO
5832	<WA>	<U30EE>	KATAKANA LETTER SMALL WA
5833	<Wa>	<U30EF>	KATAKANA LETTER WA
5834	<Wi>	<U30F0>	KATAKANA LETTER WI
5835	<We>	<U30F1>	KATAKANA LETTER WE
5836	<Wo>	<U30F2>	KATAKANA LETTER WO
5837	<N6>	<U30F3>	KATAKANA LETTER N
5838	<Vu>	<U30F4>	KATAKANA LETTER VU
5839	<KA>	<U30F5>	KATAKANA LETTER SMALL KA
5840	<KE>	<U30F6>	KATAKANA LETTER SMALL KE
5841	<Va>	<U30F7>	KATAKANA LETTER VA
5842	<Vi>	<U30F8>	KATAKANA LETTER VI
5843	<Ve>	<U30F9>	KATAKANA LETTER VE
5844	<Vo>	<U30FA>	KATAKANA LETTER VO
5845	<. 6>	<U30FB>	KATAKANA MIDDLE DOT
5846	<- 6>	<U30FC>	KATAKANA-HIRAGANA PROLONGED SOUND MARK
5847	<* 6>	<U30FD>	KATAKANA ITERATION MARK
5848	<+ 6>	<U30FE>	KATAKANA VOICED ITERATION MARK
5849	<b4>	<U3105>	BOPOMOFO LETTER B
5850	<p4>	<U3106>	BOPOMOFO LETTER P
5851	<m4>	<U3107>	BOPOMOFO LETTER M
5852	<f4>	<U3108>	BOPOMOFO LETTER F
5853	<d4>	<U3109>	BOPOMOFO LETTER D
5854	<t4>	<U310A>	BOPOMOFO LETTER T
5855	<n4>	<U310B>	BOPOMOFO LETTER N
5856	<l4>	<U310C>	BOPOMOFO LETTER L
5857	<g4>	<U310D>	BOPOMOFO LETTER G
5858	<k4>	<U310E>	BOPOMOFO LETTER K
5859	<h4>	<U310F>	BOPOMOFO LETTER H
5860	<j4>	<U3110>	BOPOMOFO LETTER J
5861	<q4>	<U3111>	BOPOMOFO LETTER Q
5862	<x4>	<U3112>	BOPOMOFO LETTER X
5863	<zh>	<U3113>	BOPOMOFO LETTER ZH
5864	<ch>	<U3114>	BOPOMOFO LETTER CH
5865	<sh>	<U3115>	BOPOMOFO LETTER SH
5866	<r4>	<U3116>	BOPOMOFO LETTER R
5867	<z4>	<U3117>	BOPOMOFO LETTER Z
5868	<c4>	<U3118>	BOPOMOFO LETTER C
5869	<s4>	<U3119>	BOPOMOFO LETTER S
5870	<a4>	<U311A>	BOPOMOFO LETTER A
5871	<o4>	<U311B>	BOPOMOFO LETTER O
5872	<e4>	<U311C>	BOPOMOFO LETTER E
5873	<eh4>	<U311D>	BOPOMOFO LETTER EH
5874	<ai>	<U311E>	BOPOMOFO LETTER AI
5875	<ei>	<U311F>	BOPOMOFO LETTER EI
5876	<au>	<U3120>	BOPOMOFO LETTER AU
5877	<ou>	<U3121>	BOPOMOFO LETTER OU
5878	<an>	<U3122>	BOPOMOFO LETTER AN
5879	<en>	<U3123>	BOPOMOFO LETTER EN
5880	<aN>	<U3124>	BOPOMOFO LETTER ANG
5881	<eN>	<U3125>	BOPOMOFO LETTER ENG
5882	<er>	<U3126>	BOPOMOFO LETTER ER
5883	<i4>	<U3127>	BOPOMOFO LETTER I
5884	<u4>	<U3128>	BOPOMOFO LETTER U
5885	<iu>	<U3129>	BOPOMOFO LETTER IU
5886	<v4>	<U312A>	BOPOMOFO LETTER V
5887	<nG>	<U312B>	BOPOMOFO LETTER NG
5888	<gn>	<U312C>	BOPOMOFO LETTER GN
5889	<(JU)>	<U321C>	PARENTHESES HANGUL CIEUC U
5890	<1c>	<U3220>	PARENTHESES IDEOGRAPH ONE
5891	<2c>	<U3221>	PARENTHESES IDEOGRAPH TWO
5892	<3c>	<U3222>	PARENTHESES IDEOGRAPH THREE
5893	<4c>	<U3223>	PARENTHESES IDEOGRAPH FOUR
5894	<5c>	<U3224>	PARENTHESES IDEOGRAPH FIVE
5895	<6c>	<U3225>	PARENTHESES IDEOGRAPH SIX
5896	<7c>	<U3226>	PARENTHESES IDEOGRAPH SEVEN
5897	<8c>	<U3227>	PARENTHESES IDEOGRAPH EIGHT
5898	<9c>	<U3228>	PARENTHESES IDEOGRAPH NINE
5899	<10c>	<U3229>	PARENTHESES IDEOGRAPH TEN
5900	<KSC>	<U327F>	KOREAN STANDARD SYMBOL
5901	<am>	<U33C2>	SQUARE AM
5902	<pm>	<U33D8>	SQUARE PM

5903	<ff>	<UFB00>	LATIN SMALL LIGATURE FF
5904	<fi>	<UFB01>	LATIN SMALL LIGATURE FI
5905	<fl>	<UFB02>	LATIN SMALL LIGATURE FL
5906	<ffi>	<UFB03>	LATIN SMALL LIGATURE FFI
5907	<ffl>	<UFB04>	LATIN SMALL LIGATURE FFL
5908	<st>	<UFB05>	LATIN SMALL LIGATURE LONG S T
5909	<st>	<UFB06>	LATIN SMALL LIGATURE ST
5910	<3+>	<UFE7D>	ARABIC SHADDA MEDIAL FORM
5911	<aM.>	<UFE82>	ARABIC LETTER ALEF WITH MADDA ABOVE FINAL FORM
5912	<aH.>	<UFE84>	ARABIC LETTER ALEF WITH HAMZA ABOVE FINAL FORM
5913	<aH.>	<UFE88>	ARABIC LETTER ALEF WITH HAMZA BELOW FINAL FORM
5914	<a+>	<UFE8D>	ARABIC LETTER ALEF ISOLATED FORM
5915	<a+>	<UFE8E>	ARABIC LETTER ALEF FINAL FORM
5916	<b+>	<UFE8F>	ARABIC LETTER BEH ISOLATED FORM
5917	<b+>	<UFE90>	ARABIC LETTER BEH FINAL FORM
5918	<b+>	<UFE91>	ARABIC LETTER BEH INITIAL FORM
5919	<b+>	<UFE92>	ARABIC LETTER BEH MEDIAL FORM
5920	<tm->	<UFE93>	ARABIC LETTER TEH MARBUTA ISOLATED FORM
5921	<tm.>	<UFE94>	ARABIC LETTER TEH MARBUTA FINAL FORM
5922	<t+>	<UFE95>	ARABIC LETTER TEH ISOLATED FORM
5923	<t+>	<UFE96>	ARABIC LETTER TEH FINAL FORM
5924	<t+>	<UFE97>	ARABIC LETTER TEH INITIAL FORM
5925	<t+>	<UFE98>	ARABIC LETTER TEH MEDIAL FORM
5926	<tk->	<UFE99>	ARABIC LETTER THEH ISOLATED FORM
5927	<tk.>	<UFE9A>	ARABIC LETTER THEH FINAL FORM
5928	<tk.>	<UFE9B>	ARABIC LETTER THEH INITIAL FORM
5929	<tk;>	<UFE9C>	ARABIC LETTER THEH MEDIAL FORM
5930	<g+>	<UFE9D>	ARABIC LETTER JEEM ISOLATED FORM
5931	<g+>	<UFE9E>	ARABIC LETTER JEEM FINAL FORM
5932	<g+>	<UFE9F>	ARABIC LETTER JEEM INITIAL FORM
5933	<g+>	<UFEA0>	ARABIC LETTER JEEM MEDIAL FORM
5934	<hk->	<UFEA1>	ARABIC LETTER HAH ISOLATED FORM
5935	<hk.>	<UFEA2>	ARABIC LETTER HAH FINAL FORM
5936	<hk.>	<UFEA3>	ARABIC LETTER HAH INITIAL FORM
5937	<hk;>	<UFEA4>	ARABIC LETTER HAH MEDIAL FORM
5938	<x+>	<UFEA5>	ARABIC LETTER KHAH ISOLATED FORM
5939	<x+>	<UFEA6>	ARABIC LETTER KHAH FINAL FORM
5940	<x+>	<UFEA7>	ARABIC LETTER KHAH INITIAL FORM
5941	<x+>	<UFEA8>	ARABIC LETTER KHAH MEDIAL FORM
5942	<d+>	<UFEA9>	ARABIC LETTER DAL ISOLATED FORM
5943	<d+>	<UFEAA>	ARABIC LETTER DAL FINAL FORM
5944	<dk->	<UFEAB>	ARABIC LETTER THAL ISOLATED FORM
5945	<dk.>	<UFEAC>	ARABIC LETTER THAL FINAL FORM
5946	<r+>	<UFEAD>	ARABIC LETTER REH ISOLATED FORM
5947	<r+>	<UFEAE>	ARABIC LETTER REH FINAL FORM
5948	<z+>	<UFEAF>	ARABIC LETTER ZAIN ISOLATED FORM
5949	<z+>	<UFEB0>	ARABIC LETTER ZAIN FINAL FORM
5950	<s+>	<UFEB1>	ARABIC LETTER SEEN ISOLATED FORM
5951	<s+>	<UFEB2>	ARABIC LETTER SEEN FINAL FORM
5952	<s+>	<UFEB3>	ARABIC LETTER SEEN INITIAL FORM
5953	<s+>	<UFEB4>	ARABIC LETTER SEEN MEDIAL FORM
5954	<sn->	<UFEB5>	ARABIC LETTER SHEEN ISOLATED FORM
5955	<sn.>	<UFEB6>	ARABIC LETTER SHEEN FINAL FORM
5956	<sn.>	<UFEB7>	ARABIC LETTER SHEEN INITIAL FORM
5957	<sn;>	<UFEB8>	ARABIC LETTER SHEEN MEDIAL FORM
5958	<c+>	<UFEB9>	ARABIC LETTER SAD ISOLATED FORM
5959	<c+>	<UFEBA>	ARABIC LETTER SAD FINAL FORM
5960	<c+>	<UFEBB>	ARABIC LETTER SAD INITIAL FORM
5961	<c+>	<UFEBC>	ARABIC LETTER SAD MEDIAL FORM
5962	<dd->	<UFEBD>	ARABIC LETTER DAD ISOLATED FORM
5963	<dd.>	<UFEBE>	ARABIC LETTER DAD FINAL FORM
5964	<dd.>	<UFEBF>	ARABIC LETTER DAD INITIAL FORM
5965	<dd;>	<UFEC0>	ARABIC LETTER DAD MEDIAL FORM
5966	<tj->	<UFEC1>	ARABIC LETTER TAH ISOLATED FORM
5967	<tj.>	<UFEC2>	ARABIC LETTER TAH FINAL FORM
5968	<tj.>	<UFEC3>	ARABIC LETTER TAH INITIAL FORM
5969	<tj;>	<UFEC4>	ARABIC LETTER TAH MEDIAL FORM
5970	<zH->	<UFEC5>	ARABIC LETTER ZAH ISOLATED FORM
5971	<zH.>	<UFEC6>	ARABIC LETTER ZAH FINAL FORM
5972	<zH.>	<UFEC7>	ARABIC LETTER ZAH INITIAL FORM
5973	<zH;>	<UFEC8>	ARABIC LETTER ZAH MEDIAL FORM
5974	<e+>	<UFEC9>	ARABIC LETTER AIN ISOLATED FORM
5975	<e+>	<UFECA>	ARABIC LETTER AIN FINAL FORM
5976	<e+>	<UFECB>	ARABIC LETTER AIN INITIAL FORM
5977	<e+>	<UFEC<	ARABIC LETTER AIN MEDIAL FORM
5978	<i+>	<UFECD>	ARABIC LETTER GHAIN ISOLATED FORM
5979	<i+>	<UFECE>	ARABIC LETTER GHAIN FINAL FORM
5980	<i+>	<UFECF>	ARABIC LETTER GHAIN INITIAL FORM
5981	<i+>	<UFED0>	ARABIC LETTER GHAIN MEDIAL FORM
5982	<f+>	<UFED1>	ARABIC LETTER FEH ISOLATED FORM
5983	<f+>	<UFED2>	ARABIC LETTER FEH FINAL FORM
5984	<f+>	<UFED3>	ARABIC LETTER FEH INITIAL FORM
5985	<f+>	<UFED4>	ARABIC LETTER FEH MEDIAL FORM
5986	<q+>	<UFED5>	ARABIC LETTER QAF ISOLATED FORM
5987	<q+>	<UFED6>	ARABIC LETTER QAF FINAL FORM
5988	<q+>	<UFED7>	ARABIC LETTER QAF INITIAL FORM
5989	<q+>	<UFED8>	ARABIC LETTER QAF MEDIAL FORM
5990	<k+>	<UFED9>	ARABIC LETTER KAF ISOLATED FORM
5991	<k+>	<UFEDA>	ARABIC LETTER KAF FINAL FORM

5992	<k+,>	<UFEDB>	ARABIC LETTER KAF INITIAL FORM
5993	<k+; >	<UFEDC>	ARABIC LETTER KAF MEDIAL FORM
5994	<l+>	<UFEDD>	ARABIC LETTER LAM ISOLATED FORM
5995	<l+.>	<UFEDE>	ARABIC LETTER LAM FINAL FORM
5996	<l+; >	<UFEDF>	ARABIC LETTER LAM INITIAL FORM
5997	<l+; >	<UFEE0>	ARABIC LETTER LAM MEDIAL FORM
5998	<m+>	<UFEE1>	ARABIC LETTER MEEM ISOLATED FORM
5999	<m+.>	<UFEE2>	ARABIC LETTER MEEM FINAL FORM
6000	<m+; >	<UFEE3>	ARABIC LETTER MEEM INITIAL FORM
6001	<m+; >	<UFEE4>	ARABIC LETTER MEEM MEDIAL FORM
6002	<n+>	<UFEE5>	ARABIC LETTER NOON ISOLATED FORM
6003	<n+.>	<UFEE6>	ARABIC LETTER NOON FINAL FORM
6004	<n+; >	<UFEE7>	ARABIC LETTER NOON INITIAL FORM
6005	<n+; >	<UFEE8>	ARABIC LETTER NOON MEDIAL FORM
6006	<h+>	<UFEE9>	ARABIC LETTER HEH ISOLATED FORM
6007	<h+.>	<UFEEA>	ARABIC LETTER HEH FINAL FORM
6008	<h+; >	<UFEEB>	ARABIC LETTER HEH INITIAL FORM
6009	<h+; >	<UFEEC>	ARABIC LETTER HEH MEDIAL FORM
6010	<w+>	<UFEEF>	ARABIC LETTER WAW ISOLATED FORM
6011	<w+.>	<UFEEE>	ARABIC LETTER WAW FINAL FORM
6012	<j+>	<UFEEF>	ARABIC LETTER ALEF MAKSURA ISOLATED FORM
6013	<j+.>	<UFEF0>	ARABIC LETTER ALEF MAKSURA FINAL FORM
6014	<y+>	<UFEF1>	ARABIC LETTER YEH ISOLATED FORM
6015	<y+.>	<UFEF2>	ARABIC LETTER YEH FINAL FORM
6016	<y+; >	<UFEF3>	ARABIC LETTER YEH INITIAL FORM
6017	<y+; >	<UFEF4>	ARABIC LETTER YEH MEDIAL FORM
6018	<lm>	<UFEF5>	ARABIC LIGATURE LAM WITH ALEF WITH MADDA ABOVE ISOLATED FORM
6019	<lm>	<UFEF6>	ARABIC LIGATURE LAM WITH ALEF WITH MADDA ABOVE FINAL FORM
6020	<lh>	<UFEF7>	ARABIC LIGATURE LAM WITH ALEF WITH HAMZA ABOVE ISOLATED FORM
6021	<lh>	<UFEF8>	ARABIC LIGATURE LAM WITH ALEF WITH HAMZA ABOVE FINAL FORM
6022	<lh>	<UFEF9>	ARABIC LIGATURE LAM WITH ALEF WITH HAMZA BELOW ISOLATED FORM
6023	<lh>	<UFEFA>	ARABIC LIGATURE LAM WITH ALEF WITH HAMZA BELOW FINAL FORM
6024	<la>	<UFEFB>	ARABIC LIGATURE LAM WITH ALEF ISOLATED FORM
6025	<la>	<UFEFC>	ARABIC LIGATURE LAM WITH ALEF FINAL FORM
6026	<"3>	<U80000000>	DIACRITICAL MARK UMLAUT <ISO-IR-53_C9/> (not a real character)
6027	<"1>	<U80000001>	DIACRITICAL MARK DIAERESIS WITH ACCENT <ISO-IR-70_C0/> (not a real character)
6028	character)		
6029	<"!>	<U80000002>	DIACRITICAL MARK GRAVE ACCENT <ISO-IR-103_C1/> (not a real character)
6030	character)		
6031	<"'>	<U80000003>	DIACRITICAL MARK ACUTE ACCENT <ISO-IR-103_C2/> (not a real character)
6032	character)		
6033	<"/>	<U80000004>	DIACRITICAL MARK CIRCUMFLEX ACCENT <ISO-IR-103_C3/> (not a real character)
6034	character)		
6035	<"?>	<U80000005>	DIACRITICAL MARK TILDE <ISO-IR-103_C4/> (not a real character)
6036	<"->	<U80000006>	DIACRITICAL MARK MACRON <ISO-IR-103_C5/> (not a real character)
6037	<"(>	<U80000007>	DIACRITICAL MARK BREVE <ISO-IR-103_C6/> (not a real character)
6038	<".>	<U80000008>	DIACRITICAL MARK DOT ABOVE <ISO-IR-103_C7/> (not a real character)
6039	<"::>	<U80000009>	DIACRITICAL MARK DIAERESIS <ISO-IR-103_C8/> (not a real character)
6040	<"0>	<U8000000A>	DIACRITICAL MARK RING ABOVE <ISO-IR-103_CA/> (not a real character)
6041	<">	<U8000000B>	DIACRITICAL MARK CEDILLA <ISO-IR-103_CB/> (not a real character)
6042	<"_>	<U8000000C>	DIACRITICAL MARK LOW LINE <ISO-IR-103_CC/> (not a real character)
6043	<">	<U8000000D>	DIACRITICAL MARK DOUBLE ACUTE ACCENT <ISO-IR-103_CD/> (not a real character)
6044	character)		
6045	<">	<U8000000E>	DIACRITICAL MARK OGONEK <ISO-IR-103_CE/> (not a real character)
6046	<"<>	<U8000000F>	DIACRITICAL MARK CARON <ISO-IR-103_CF/> (not a real character)
6047	<"=>	<U80000010>	DIACRITICAL MARK DOUBLE LOW LINE <ISO-IR-38_D9/> (not a real character)
6048	character)		
6049	<"//>	<U80000011>	DIACRITICAL MARK LONG SOLIDUS OVERLAY <ISO-IR-128_C9/> (not a real character)
6050	character)		
6051	<"p>	<U80000012>	GREEK DIACRITICAL MARK PSILI PNEUMATA <ISO-IR-55_25/> (not a real character)
6052	character)		
6053	<"d>	<U80000013>	GREEK DIACRITICAL MARK DASIA PNEUMATA <ISO-IR-55_26/> (not a real character)
6054	character)		
6055	<"i>	<U80000014>	GREEK DIACRITICAL MARK IOTA BELOW <ISO-IR-55_27/> (not a real character)
6056	character)		
6057	<+>	<U80000015>	IDEOGRAPHIC DITTO MARK <ISO-IR-87_2138/>
6058	<a+; >	<U80000016>	ARABIC LETTER ALEF FINAL FORM COMPATIBILITY <IBM868_90/>
6059	<Tel>	<U80000017>	TEL COMPATIBILITY SIGN <ISO-IR-149_2265/>
6060	<UA>	<U80000018>	Unit space A <ISO-IR-8-1_40/>
6061	<UB>	<U80000019>	Unit space B <ISO-IR-8-1_60/>
6062	<t3>	<U8000001A>	GREEK SMALL LETTER STIGMA <ISO-IR-55_47/>
6063	<m3>	<U8000001B>	GREEK SMALL LETTER DIGAMMA <ISO-IR-55_48/>
6064	<k3>	<U8000001C>	GREEK SMALL LETTER KOPPA <ISO-IR-55_54/>
6065	<p3>	<U8000001D>	GREEK SMALL LETTER SAMPI <ISO-IR-55_5E/>
6066	<Mc>	<U8000001E>	APPLE LOGO (Macintosh_F0)
6067	<Fl>	<U8000001F>	HUNGARIAN FLORINTH (CWI_9F)
6068	<Ss>	<U80000020>	LATIN CAPITAL LIGATURE SS (German) (CORK_FF)
6069	<Ch>	<U80000021>	LATIN SMALL LIGATURE CH (Slovak) (KOI-8_CS2_C7)
6070	<CH>	<U80000022>	LATIN CAPITAL LIGATURE CH (Slovak) (KOI-8_CS2_E7)
6071	<//c>	<U80000024>	JOIN THIS LINE WITH NEXT LINE (Mnemonic)
6072	<H->	<U0023>	NUMBER SIGN
6073	<!S>	<U0024>	DOLLAR SIGN
6074	<@>	<U0040>	COMMERCIAL AT
6075	<Oa>	<U0040>	COMMERCIAL AT
6076	<!C>	<U00A2>	CENT SIGN
6077	<L->	<U00A3>	POUND SIGN
6078	<Xo>	<U00A4>	CURRENCY SIGN

6079	<Y->	<U00A5>	YEN SIGN
6080	<!B>	<U00A6>	BROKEN BAR
6081	<S>	<U00A7>	SECTION SIGN
6082	<OC>	<U00A9>	COPYRIGHT SIGN
6083	<7!>	<U00AC>	NOT SIGN
6084	<OR>	<U00AE>	REGISTERED SIGN
6085	<9I>	<U00B6>	PILCROW SIGN
6086	<_>	<U2500>	BOX DRAWINGS LIGHT HORIZONTAL
6087	<_=>	<U2501>	BOX DRAWINGS HEAVY HORIZONTAL
6088	<_!>	<U2502>	BOX DRAWINGS LIGHT VERTICAL
6089	<_V/>	<U250C>	BOX DRAWINGS LIGHT DOWN AND RIGHT
6090	<_V<w>	<U2510>	BOX DRAWINGS LIGHT DOWN AND LEFT
6091	<_A/>	<U2514>	BOX DRAWINGS LIGHT UP AND RIGHT
6092	<_A<>	<U2518>	BOX DRAWINGS LIGHT UP AND LEFT
6093	<_!/>	<U251C>	BOX DRAWINGS LIGHT VERTICAL AND RIGHT
6094	<_!<>	<U2524>	BOX DRAWINGS LIGHT VERTICAL AND LEFT
6095	<_V->	<U252C>	BOX DRAWINGS LIGHT DOWN AND HORIZONTAL
6096	<_A>	<U2534>	BOX DRAWINGS LIGHT UP AND HORIZONTAL
6097	<_!->	<U253C>	BOX DRAWINGS LIGHT VERTICAL AND HORIZONTAL
6098	<_V/>/>	<U2571>	BOX DRAWINGS LIGHT DIAGONAL UPPER RIGHT TO LOWER LEFT
6099	<_<>	<U2572>	BOX DRAWINGS LIGHT DIAGONAL UPPER LEFT TO LOWER RIGHT
6100	<_./>///>	<U25E2>	BLACK LOWER RIGHT TRIANGLE
6101	<_.<>	<U25E3>	BLACK LOWER LEFT TRIANGLE
6102	<_d!>	<U266A>	EIGHTH NOTE

6104

## 7 CONFORMANCE

6105

6106

6107

### 7.1 FDCC-set

6108

A FDCC-set description is conforming to this standard if it meets the requirements in clause 4.

6110

6111

6112

### 7.2 FDCC-set category

6113

Conformance can be claimed for a category description against each of the clauses 4.2 thru 4.11, and then the requirements of clause 4.0 shall also be met, and a LC\_IDENTIFICATION category as described in clause 4.1 shall be specified.

6116

6117

6118

### 7.3 Charmap

6119

A charmap description is conforming to this standard if it meets the requirements in clause

6120

6121

6122

### 7.4 Repertoiremap

6123

6124

6125

A repertoiremap description is conforming to this standard if it meets the requirements in

6126

6127

clause 6.

6128  
6129  
6130  
6131  
6132  
6133  
6134  
6135  
6136  
6137  
6138  
6139  
6140  
6141  
6142  
6143  
6144  
6145  
6146  
6147  
6148  
6149  
6150  
6151  
6152  
6153  
6154  
6155  
6156  
6157  
6158  
6159  
6160  
6161  
6162  
6163  
6164  
6165  
6166  
6167  
6168  
6169  
6170  
6171  
6172  
6173  
6174  
6175  
6176  
6177

## Annex A (informative)

### Differences from the ISO/IEC 9945-2 standard

This standard originated from the locale and charmap specifications in the ISO/IEC 9945-2 standard, and it intends to be backwards compatible, so that what is conformant to that standard should also be conformant to this standard.

A number of enhancements have been done and a number of restrictions have been lifted in comparison to the POSIX standard:

#### A.1 Restrictions removed

1. Dependence on specific meaning of the character NUL as termination of a string (from the C standard) has been removed, to cater for other programming languages than C.

#### A.2 Enhancements

1. A description of a "repertoiremap" definition was added to facilitate descriptions of FDCC-sets without charmaps, and also to provide binding from a FDCC-set using one set of character names to charmaps using another naming set.

2. The specific POSIX locale has been replaced with the "i18n" FDCC-set, defined on the repertoire on ISO/IEC 10646.

3. Transliteration support has been added in the LC\_CTYPE category.

4. Terminology has been aligned with ISO/IEC TR 11017, especially the POSIX term "locale" has been changed to "FDCC-set".

5. A date escape format "%F" has been added for ISO 8601 dates, and another date escape format "%f" has been added for weekday number with Monday being the first day of the week.

6. Added to LC\_MONETARY to accommodate differences between local and international formats:

```
int_p_cs_precedes
int_p_sep_by_space
int_n_cs_precedes
int_n_sep_by_space
```

7. Section symbols have been added via the "section-symbol" keyword in the LC\_COLLATE category.

8. The "order\_start" keyword has got an optional "section-symbol" identifier

9. The keywords "reorder-sections-after" and "reorder-sections\_end" have been introduced to reorder sections.



- 6178 10. Symbolic elipses (both decimal and hexadecimal) has been introduced as a notation.  
6179
- 6180 11. The "print" CTYPE class includes automatically all "graph" characters.  
6181
- 6182 12. The <Uxxxx> and <Uxxxxxxxx> notations have been introduced as predefined  
6183 symbolic character names, together with a number of symbolic character names derived  
6184 from POSIX and the Internet.  
6185
- 6186 13. Toggling commands define, undef, ifdef, ifndef, elif, else, and endif have been  
6187 introduced for the FDCC-set category LC\_COLLATE, in the style of the C-precompiler.  
6188
- 6189 14. New categories LC\_IDENTIFICATION, LC\_PAPER, LC\_NAME, LC\_ADDRESS,  
6190 and LC\_TELEPHONE, have been introduced.  
6191
- 6192 15. The LC\_CTYPE has got support for new classes, via the new keywords class and  
6193 map, which corresponds to the C standard library functions iswctype() and towctrans()  
6194 respectively.  
6195
- 6196 16. The "digit" keyword now supports digits for multiple scripts.  
6197
- 6198 17. The LC\_MONETARY category provides support for multiple currencies, such as the  
6199 native currency and the Euro in some European countries.  
6200
- 6201 18. The LC\_TIME has got a number of enhancements to cater for alternate calendars, and  
6202 timezone information may be given.  
6203
- 6204 19. The charmap specification has been enhanced to support ISO 2022.

6205  
6206  
6207  
6208  
6209  
6210  
6211  
6212  
6213  
6214  
6215  
6216  
6217  
6218  
6219  
6220  
6221  
6222  
6223  
6224  
6225  
6226  
6227  
6228  
6229  
6230  
6231  
6232  
6233  
6234  
6235  
6236  
6237  
6238  
6239  
6240  
6241  
6242  
6243  
6244  
6245  
6246  
6247  
6248  
6249  
6250  
6251  
6252  
6253  
6254

## **Annex B** (informative)

### **Rationale**

#### **B.1 FDCC-set Rationale**

The description of FDCC-sets is based on work performed in the UniForum Technical Committee Subcommittee on Internationalisation and on POSIX. Wherever appropriate, keywords were taken from the C Standard or the POSIX-2 standard. The C and POSIX term "locale" has been changed into the term "FDCC-set" from ISO/IEC TR 11017 to align with that specification.

The POSIX utility "localedef" compiles locale sources into object files. The "object" definitions need not be portable, as long as "source" definitions are. Strictly speaking, "source" definitions are portable only between applications using the same character set(s). Such "source" definitions can, if they use symbolic names only, easily be ported between systems using different code sets as long as the characters in the portable character set (ISO 646) have common values between the code sets; this is frequently the case in historical applications. Of course, this requires that the symbolic names used for characters outside the portable character set are identical between character sets.

To avoid confusion between an octal constant and a backreference, the octal, hexadecimal, and decimal constants must contain at least two digits. As single-digit constants are relatively rare, this should not impose any significant hardship. Each of the constants includes "two or more" digits to account for systems in which the byte size is larger than eight bits. For example, an ISO/IEC 10646 system that has defined 16-bit bytes may require six octal, four hexadecimal, and five decimal digits, for some coded characters.

As an international (ISO/IEC) standard this standard should follow the ISO/IEC guidelines, including the ISO/IEC TR 10176. This TR has a rule that characters outside the invariant part of ISO/IEC 646 should not be used in portable specifications. The backslash and the number-sign character are not in the invariant part. As far as general usage of these symbols, they are covered by the "grandfather clause" specifying previous practise in international standards and in the industry such as in specifications from The Open Group, but for newly defined interfaces, ISO has requested that specifications provide alternate representations, and this standard then follows POSIX for backward compatibility. Consequently, while the default escape character remains the backslash, and the default comment character is the number-sign, applications are required to recognize alternative representations, identified in the applicable source text via the "escape\_char" and "comment\_char" keywords.

##### **B.1.1 LC\_IDENTIFICATION Rationale.**

The LC\_IDENTIFICATION category gives meta-information on the FDCC-set, such as who created it, and what is the level of conformance for each of the FDCC sets.

### 6255 **B.1.2 LC\_CTYPE Rationale**

6256

6257 The LC\_CTYPE category primarily is used to define the encoding-independent aspects of  
6258 a character set, such as character classification. In addition, certain encoding-dependent  
6259 characteristics are also defined for an application via the LC\_CTYPE category. This  
6260 standard does not mandate that the encoding used in the FDCC-set is the same as the one  
6261 used by the application, because an application may decide that it is advantageous to  
6262 define a FDCC-set in a system-wide encoding rather than having multiple, logically  
6263 identical FDCC-sets in different encodings, and to convert from the application encoding  
6264 to the system-wide encoding on usage. Other applications could require encoding-depen-  
6265 dent FDCC-sets. In either case, the LC\_CTYPE attributes that are directly dependent on  
6266 the encoding, such as "mb\_cur\_max" and the display width of characters, are not user-  
6267 specifiable in a locale source, and are consequently not defined as keywords.

6268

6269 As the LC\_CTYPE character classes are based on the C Standard character-class  
6270 definition, the category does not support multicharacter elements. For instance, the  
6271 German character <sharp-s> is traditionally classified as a lowercase letter. There is no  
6272 corresponding uppercase letter; in proper capitalization of German text the <sharp-s> will  
6273 be replaced by SS; i.e., by two characters. This kind of conversion is outside the scope of  
6274 the "toupper" and "tolower" keywords.

6275

6276 The character classes "digit", "xdigit", "lower", "upper", and "space" have a set of  
6277 automatically included characters. These only need to be specified if the character values  
6278 (i.e. encoding) differs from the application default values. The definition of character class  
6279 "digit" allows alternate digits (e.g., Hindi) to be specified here. The definition of character  
6280 class "xdigit" requires that the characters included in character class "digit" are included  
6281 here also, and allows for different symbols for the hexadecimal digits 10 through 15.

6282

6283 The "combining" and "combining-level3" classes are an IT-enablement of ISO/IEC 10646  
6284 definitions of combining characters. These can be used to check identifiers for consistence  
6285 with the guidelines given in TR 10176 annex A.

6286

6287

### 6288 **B.1.3 LC\_COLLATE Rationale.**

6289

6290 The LC\_COLLATE category governs the collation order in the FDCC-set, and may thus  
6291 be useful for the processing of the ISO/IEC 14651 string ordering and comparison  
6292 standard, the C Standard strxfrm() and strcoll() functions, as well as a number of POSIX-2  
6293 utilities.

6294

6295 The rules governing collation depends to some extent on the use. At least five different  
6296 levels of increasingly complex collation rules can be distinguished:

6297

- 6298 (1) Byte/machine code order. This is the historical collation order in the UNIX  
6299 system and many proprietary operating systems. Collation is here done  
6300 character by character, without any regard to context. The primary virtue is that  
6301 it usually is quite fast, and also completely deterministic; it works well when  
6302 the native machine collation sequence matches the user expectations.
- 6303 (2) Character order. On this level, collation is also done character by character,  
6304 without regard to context. The order between characters is, however, not deter-

- 6305           mined by the code values, but on the user's expectations of the correct order  
 6306           between characters. In addition, such a (simple) collation order can specify that  
 6307           certain characters collate equal (e.g., upper and lowercase letters).
- 6308   (3)       String ordering. On this level, entire strings are compared based on relatively  
 6309           straightforward rules. At this level, several "passes" may be required to deter-  
 6310           mine the order between two strings. Characters may be ignored in some passes,  
 6311           but not in others; the strings may be compared in different directions; and  
 6312           simple string substitutions may be made before strings are compared. This level  
 6313           is best described as "dictionary" ordering; it is based on the spelling, not the  
 6314           pronunciation, or meaning, of the words.
- 6315   (4)       Text search ordering. This is a further refinement of the previous level, best de-  
 6316           scribed as "telephone book ordering"; some common homonyms (words spelled  
 6317           differently but with same pronunciation) are collated together; numbers are  
 6318           collated as if spelled with words, and so on.
- 6319   (5)       Semantic level ordering. Words and strings are collated based on their meaning;  
 6320           entire words (such as "the") are eliminated, the ordering is not deterministic.  
 6321           This may requires special software, and is highly dependent on the intended  
 6322           use.

6323  
 6324   While the historical collation order formally is at level 1, for the English language it  
 6325   corresponds roughly to elements at level 2. The user expects to see the output from the  
 6326   "ls" utility sorted very much as it would be in a dictionary. While telephone book ordering  
 6327   would be an optimal goal for standard collation, this was ruled out as the order would be  
 6328   language dependent. Furthermore, a requirement was that the order must be determined  
 6329   solely from the text string and the collation rules; no external information (e.g., "pronu-  
 6330   nciation dictionaries") could be required.

6331  
 6332   As a result, the goal for the collation support is at level 3. This also matches the re-  
 6333   quirements for the Canadian collation order standard, as well as other, known collation  
 6334   requirements for alphabetic scripts. It specifically rules out collation based on pronun-  
 6335   ciation rules, or based on semantic analysis of the text. The syntax for the LC\_COLLATE  
 6336   category source is the result of a cooperative effort between representatives for many  
 6337   countries and organizations working with international issues, such as UniForum, X/Open,  
 6338   and ISO, and it meets the requirements for level 3, and has been verified to produce the  
 6339   correct result with examples based on Canadian and Danish collation order.

6340  
 6341   The directives that can be specified in an operand to the order\_start keyword are based on  
 6342   the requirements specified in several proposed standards and in customary use. The  
 6343   following is a rephrasing of rules defined for "lexical ordering in English and French" by  
 6344   the Canadian Standards Association (text in brackets is rephrased):

- 6345  
 6346   (1)       Once special characters (punctuation) have been removed from original strings,  
 6347           the ordering is determined by scanning forward (left to right) [disregarding case  
 6348           and diacriticals].
- 6349   (2)       In case of equivalence, special characters are once again removed from original  
 6350           strings and the ordering is determined scanning backward (starting from the  
 6351           rightmost character of the string and back), character by character, (disregarding  
 6352           case but considering diacriticals).
- 6353   (3)       In case of repeated equivalence, special characters are removed again from  
 6354           original strings and the ordering is determined scanning forward, character by

6355 character, (considering both case and diacriticals).  
6356 (4) If there is still an ordering equivalence after rules (1) through (3) have been  
6357 applied, then only special characters and the position they occupy in the string  
6358 are considered to determine ordering. The string that has a special character in  
6359 the lowest position comes first. If two strings have a special character in the  
6360 same position, the character [with the lowest collation value] comes first. In  
6361 case of equality, the other special characters are considered until there is a  
6362 difference or all special characters have been exhausted.  
6363

6364 It is estimated that the standard covers the requirements for all European languages, and  
6365 no particular problems are anticipated for Cyrillic or Middle Eastern scripts.  
6366

6367 The Far East (particularly Japanese/Chinese) collations are often based on contextual  
6368 information. In Japan, collations of strings containing CJK characters (ideograms) are  
6369 often done considering some related information such as pronunciation, which needs a  
6370 bulk dictionary (and some common sense). Such collation, in general, falls outside the  
6371 desired goal of the standard, and the standard can support only a restricted of collations  
6372 used in Japan. There are, however, several other collation rules (stroke/radical, or "most  
6373 common pronunciation") which can be supported with the mechanism described here.  
6374 Previous drafts contained a substitute statement, which performed a regular expression  
6375 style replacement before string compares. It has been withdrawn based on balloter  
6376 objections that it was not required for the types of ordering this standard is aimed at.  
6377

6378 The character (and collating element) order is defined by the order in which characters and  
6379 elements are specified between the order\_start and order\_end keywords. This character  
6380 order is used in range expressions in regular expressions. Weights assigned to the charac-  
6381 ters and elements define the collation sequence; in the absence of weights, the character  
6382 order is also the collation sequence.  
6383

6384 The position keyword was introduced to provide the capability to consider, in a compare,  
6385 the relative position of non-IGNORED characters. As an example, consider the two strings  
6386 "o-ring" and "or-ing". Assuming the hyphen is IGNORED on the first pass, the two strings  
6387 will compare equal, and the position of the hyphen is immaterial. On second pass, all  
6388 characters except the hyphen are IGNORED, and in the normal case the two strings would  
6389 again compare equal. By taking position into account, the first collates before the second.  
6390

### 6391 **B.1.3.1 "reorder-after" rationale**

6392

6393 Much work has been done on FDCC-sets, making them quite general. The POSIX-2  
6394 standard introduced a "copy" command for all categories of the POSIX locale. This is  
6395 useful for many purposes and it ensures that two FDCC-sets are equivalent for this  
6396 category. A further step in building on previous FDCC-set work is defined in this stan-  
6397 dard.  
6398

6399 Collating sequences often vary a bit from country to country, and from language to  
6400 language, but generally much of the collating sequence is the same. For example the  
6401 Danish sequence is for the most part the same as the German or English collation, but for  
6402 about a dozen letters it differs. The same can be said for Swedish or Hungarian: generally  
6403 the Latin collating sequence is the same, but a few characters are different.  
6404

6405 This standard defines a FDCC-set defined on the character repertoire of the ISO/IEC  
6406 10646 standard, in a character set independent way. The intention is that some of the  
6407 information from this FDCC-set will be acceptable in many cultures, and that it can serve  
6408 as the basis for modifications in other cultures, to obtain a culturally acceptable  
6409 specification. Using the "reorder-after" construct will also help improve the overview of  
6410 what the changes really are for implementers and other users.

6411  
6412 An example of the use of the "reorder-after" construct is the following. A default  
6413 international ordering for the Latin alphabet may be adequate for Danish, with the  
6414 exception of the collation rules for the letters Û, ü, Æ, æ, Ä, ä, Ø, ø, Ö, ö, Å and å. By  
6415 applying the "reorder-after" construct, the Danish specification can be made more easily  
6416 by copying and reordering the existing international specification, rather than specifying  
6417 collation parameters for all Latin letters (with or without diacritics). There is no obligation  
6418 for Denmark to take this approach, but the "reorder-after" construct provides the  
6419 mechanism for doing so if it is deemed desirable.

6420  
6421  
6422  
6423  
6424  
6425  
6426  
6427  
6428  
6429  
6430  
6431  
6432  
6433  
6434  
6435  
6436  
6437  
6438  
6439  
6440  
6441  
6442  
6443  
6444  
6445  
6446  
6447  
6448  
6449  
6450  
6451  
6452  
6453  
6454  
6455  
6456  
6457  
6458  
6459  
6460  
6461  
6462  
6463  
6464  
6465  
6466  
6467  
6468  
6469  
6470  
6471  
6472

### B.1.3.2 awk script for "reorder-after" construct

A script has been written in the "awk" language defined in the POSIX standard ISO/IEC 9945-2 to implement the "reorder-after" construct. It functions as follows: It reads all of the FDCC-set and if in the LC\_COLLATE category, it processes the line, else it just outputs the line. For the LC\_COLLATE category it reads the lines and puts it into a double linked list of strings identified by a line number; at the end of the LC\_COLLATE category all the lines are output. If the line is a "copy" keyword and it reads the file referenced, extracting the LC\_COLLATE section of the file in to the list of strings. If the line is a "reorder-after" keyword, it sets a pointer to be the line number of the symbol to of the "reorder-after" keyword. If the line is part of the "reorder-after" specification, it is entered into the double linked list at this point, and the previous entry in the double linked list for the <collation-element> is removed from the list. A "reorder-end" keyword terminates the reordering.

```
BEGIN { comment = "%"; back[0]= follow[0] = 0; }
/LC_COLLATE/ { coll=1 }
/END LC_COLLATE/ { coll=0; for (lnr= 1; lnr; lnr= follow[lnr]) print cont[lnr] }

{ if (coll == 0) print $0 ;
  else { if ($1 == "copy") {
        file = $2
        while (getline < file )
        if ( $1 == "LC_COLLATE" ) copy_lc = 1
        else if ( $1 == "END" && $2 == "LC_COLLATE" ) copy_lc =0
        else if (copy_lc) {
            lnr++
            follow[lnr-1] = lnr; back [ lnr ] = lnr-1
            cont[lnr] = $0; symb[ $1 ] = lnr
        }
        close (file )
    }
    else if ($1 == "reorder-after") { ra=1 ; after = symb [ $2 ] }
    else if ($1 == "reorder-end") ra = 0
    else {
        lnr++
        if (ra) follow [ lnr ] = follow [ after ]
        if (ra) back [ follow [ after ] ] = lnr
        follow[after] = lnr; back [ lnr ] = after
        cont[lnr] = $0
        if ( ra && $1 != comment && $1 != " ) {
            old = symb [ $1 ];
            follow [ back [ old ] ] = follow [ old ];
            back [ follow [ old ] ] = back [ old ];
            symb[ $1 ] = lnr;
        }
        after = lnr
    }
  }
}
```

```

6473 B.1.3.3 Sample FDCC-set specification for Danish
6474
6475 escape_char /
6476 comment_char %
6477 repertoiremap "i18nrep"
6478 charset "ISO_8859-1:1987"
6479 % Distribution and use is free, also
6480 % for commercial purposes.
6481
6482 LC_VERSION
6483 title "Danish language FDCC-set for Denmark"
6484 source "Danish Standards Association"
6485 address "Kollegievej 6, DK-2920 Charlottenlund, Danmark"
6486 contact "Keld Simonsen"
6487 email "Keld.Simonsen@dkuug.dk"
6488 tel "+45 - 3996-6101"
6489 fax "+45 - 3996-6202"
6490 language "da"
6491 territory "DK"
6492 revision "4.2"
6493 date "1997-12-22"
6494
6495 category i18n:1998;LC_IDENTIFICATION
6496 category i18n:1998;LC_CTYPE
6497 category i18n:1998;LC_COLLATE
6498 category i18n:1998;LC_TIME
6499 category posix:1993;LC_NUMERIC
6500 category i18n:1998;LC_MONETARY
6501 category posix:1993;LC_MESSAGES
6502 category i18n:1998;LC_PAPER
6503 category i18n:1998;LC_NAME
6504 category i18n:1998;LC_ADDRESS
6505 category i18n:1998;LC_TELEPHONE
6506
6507 END LC_VERSION
6508
6509 LC_CTYPE
6510 copy "i18n"
6511 END LC_CTYPE
6512
6513 LC_COLLATE
6514 % The ordering algorithm is in accordance
6515 % with Danish Standard DS 377 (1980)
6516 % and the Danish Orthography Dictionary
6517 % (Retskrivningsordbogen, 2. udgave, 1996).
6518 % It is also in accordance with
6519 % Greenlandic orthography.
6520
6521 collating-element <A-A> from "<A><A>"
6522 collating-element <A-a> from "<A><a>"
6523 collating-element <a-A> from "<a><A>"
6524 collating-element <a-a> from "<a><a>"
6525 copy i18n
6526 reorder-after <CAPITAL>
6527 <CAPITAL>
6528 <CAPITAL-SMALL>
6529 <SMALL-CAPITAL>
6530 <SMALL>
6531 reorder-after <q8>
6532 <kk> <Q>;<SPECIAL>;<SMALL>;IGNORE
6533 reorder-after <t8>
6534 <TH> "<T><H>";"<TH><TH>";"<CAPITAL><CAPITAL>";IGNORE
6535 <th> "<T><H>";"<TH><TH>";"<SMALL><SMALL>";IGNORE
6536 reorder-after <y8>
6537 % <U:> and <U"> are treated as <Y> in Danish
6538 <U:> <Y>;<U:>;<CAPITAL>;IGNORE
6539 <u:> <Y>;<U:>;<SMALL>;IGNORE
6540 <U"> <Y>;<U">;<CAPITAL>;IGNORE
6541 <u"> <Y>;<U">;<SMALL>;IGNORE

```



```

6542 reorder-after <z8>
6543 % <AE> is a separate letter in Danish
6544 <AE> <AE>;<NONE>;<CAPITAL>;IGNORE
6545 <ae> <AE>;<NONE>;<SMALL>;IGNORE
6546 <AE'> <AE>;<ACUTE>;<CAPITAL>;IGNORE
6547 <ae'> <AE>;<ACUTE>;<SMALL>;IGNORE
6548 <A3> <AE>;<MACRON>;<CAPITAL>;IGNORE
6549 <a3> <AE>;<MACRON>;<SMALL>;IGNORE
6550 <A:> <AE>;<SPECIAL>;<CAPITAL>;IGNORE
6551 <a:> <AE>;<SPECIAL>;<SMALL>;IGNORE
6552 % <O//> is a separate letter in Danish
6553 <O//> <O//>;<NONE>;<CAPITAL>;IGNORE
6554 <o//> <O//>;<NONE>;<SMALL>;IGNORE
6555 <O//'> <O//>;<ACUTE>;<CAPITAL>;IGNORE
6556 <o//'> <O//>;<ACUTE>;<SMALL>;IGNORE
6557 <O:> <O//>;<DIAERESIS>;<CAPITAL>;IGNORE
6558 <o:> <O//>;<DIAERESIS>;<SMALL>;IGNORE
6559 <O"> <O//>;<DOUBLE-ACUTE>;<CAPITAL>;IGNORE
6560 <o"> <O//>;<DOUBLE-ACUTE>;<SMALL>;IGNORE
6561 % <AA> is a separate letter in Danish
6562 <AA> <AA>;<NONE>;<CAPITAL>;IGNORE
6563 <aa> <AA>;<NONE>;<SMALL>;IGNORE
6564 <A-A> <AA>;<A-A>;<CAPITAL>;IGNORE
6565 <A-a> <AA>;<A-A>;<CAPITAL-SMALL>;IGNORE
6566 <a-A> <AA>;<A-A>;<SMALL-CAPITAL>;IGNORE
6567 <a-a> <AA>;<A-A>;<SMALL>;IGNORE
6568 <AA'> <AA>;<AA'>;<CAPITAL>;IGNORE
6569 <aa'> <AA>;<AA'>;<SMALL>;IGNORE
6570 reorder-end
6571 END LC_COLLATE
6572
6573 LC_MONETARY
6574 int_curr_symbol " <D><K><K><SP> "
6575 currency_symbol " <k><r> "
6576 mon_decimal_point " <,> "
6577 mon_thousands_sep " <.> "
6578 mon_grouping 3;3
6579 positive_sign " "
6580 negative_sign " <-> "
6581 int_frac_digits 2
6582 frac_digits 2
6583 p_cs_precedes 1
6584 p_sep_by_space 2
6585 n_cs_precedes 1
6586 n_sep_by_space 2
6587 p_sign_posn 4
6588 n_sign_posn 4
6589 END LC_MONETARY
6590
6591 LC_NUMERIC
6592 decimal_point " <,> "
6593 thousands_sep " <.> "
6594 grouping 3;3
6595 END LC_NUMERIC
6596
6597 LC_TIME
6598 abday " <m><a><n> " ; /
6599 " <t><i><r> " ; " <o><n><s> " ; /
6600 " <t><o><r> " ; " <f><r><e> " ; /
6601 " <l><o//><r> " ; " <s><o/><n>
6602 day " <m><a><n><d><a><g> " ; /
6603 " <t><i><r><s><d><a><g> " ; /
6604 " <o><n><s><d><a><g> " ; /
6605 " <t><o><r><s><d><a><g> " ; /
6606 " <f><r><e><d><a><g> " ; /
6607 " <l><o//><r><d><a><g> " /
6608 " <s><o//><n><d><a><g> " ;
6609 week 7;19971201;4
6610 abmon " <j><a><n> " ; " <f><e><b> " ; /
6611 " <m><a><r> " ; " <a><p><r> " ; /
6612 " <m><a><j> " ; " <j><u><n> " ; /

```

```

6613         "<j><u><l>" ; "<a><u><g>" ; /
6614         "<s><e><p>" ; "<o><k><t>" ; /
6615         "<n><o><v>" ; "<d><e><c>"
6616 mon      "<j><a><n><u><a><r>" ; /
6617         "<f><e><b><r><u><a><r>" ; /
6618         "<m><a><r><t><s>" ; /
6619         "<a><p><r><i><l>" ; /
6620         "<m><a><j>" ; /
6621         "<j><u><n><i>" ; /
6622         "<j><u><l><i>" ; /
6623         "<a><u><g><u><s><t>" ; /
6624         "<s><e><p><t><e><m><b><e><r>" ; /
6625         "<o><k><t><o><b><e><r>" ; /
6626         "<n><o><v><e><m><b><e><r>" ; /
6627         "<d><e><c><e><m><b><e><r>"
6628 d_t_fmt   "<%><a><SP><%><F><SP><%><T><SP><%><Z>"
6629 d_fmt     "<%><O><d><.><SP><%><B><SP><%><Y>"
6630 atl_digits "<0><.>;<1><.>;<2><.>;<3><.>;<4><.>; /
6631         <5><.>;<6><.>;<7><.>;<8><.>;<9><.>; /
6632         <1><0><.>;<1><1><.>;<1><2><.>;<1><3><.>;<1><4><.>; /
6633         <1><5><.>;<1><6><.>;<1><7><.>;<1><8><.>;<1><9><.>; /
6634         <2><0><.>;<2><1><.>;<2><2><.>;<2><3><.>;<2><4><.>; /
6635         <2><5><.>;<2><6><.>;<2><7><.>;<2><8><.>;<2><9><.>; /
6636         <3><0><.>;<3><1><.>"
6637 t_fmt     "<%><T>"
6638 am_pm    "" ; ""
6639 t_fmt_ampm ""
6640 timezone "<C><E><T><-><1><C><E><T><SP><D><S><T><,><M><3><.><5><.><0>/
6641         <,><M><1><0><.><5><.><0>"
6642 END LC_TIME
6643
6644 LC_MESSAGES
6645 yesexpr  "<<(><1><J><j><Y><y><)/>><.><*>"
6646 noexpr   "<<(><0><N><n><)/>><.><*>"
6647 END LC_MESSAGES
6648
6649 LC_PAPER
6650 copy "i18n"
6651 END LC_PAPER
6652
6653 LC_NAME
6654 name_fmt "<%><p><%><t><%><g><%><t><%><m><%><t><%><f>"
6655 name_gen ""
6656 name_mr  "<h><r>"
6657 name_mrs "<f><r><u>"
6658 name_miss "<f><r><o></><k><e><n>"
6659 name_ms  "<f><r>"
6660 END LC_NAME
6661
6662 LC_ADDRESS
6663 country_name "<D><a><n><m><a><r><k>"
6664 country_post "<D><K>"
6665 country_ab2 "<D><K>"
6666 country_ab3 "<D><N><K>"
6667 country_num 208
6668 country_car "<D><K>"
6669 country_isbn "<8><7>"
6670 lang_ab     "<d><a>"
6671 lang_term   "<d><a><n>"
6672 postal_fmt  "<%><a><%><N><%><f><%><N><%><d><%><N><%><b><%><N><%>/

```

```
6673      <%><s><SP><%><h><SP><%><e><SP><%><r><%><N>/
6674      <%><C><-><%><z><SP><%><T><%><N><%><c><%><N>"
```

```
6675  END LC_ADDRESS
```

```
6676
```

```
6677  LC_TELEPHONE
```

```
6678  tel_int_fmt    "<+><%><c><SP><%><a><SP><%><l>"
```

```
6679  tel_dom_fmt    "<%><l>"
```

```
6680  int_select     "<0><0>"
```

```
6681  int_prefix     "<4><5>"
```

```
6682  END LC_TELEPHONE
```

```
6683
```

#### 6684 B.1.4 LC\_MONETARY Rationale.

```
6685
```

6686 The currency symbol does not appear in LC\_MONETARY because it is not defined in the  
6687 C Standard's C locale. The C Standard limits the size of decimal points and thousands  
6688 delimiters to single-byte values. In FDCC-sets based on multibyte coded character sets this  
6689 cannot be enforced, obviously; this standard does not prohibit such characters, but makes  
6690 the behaviour unspecified (in the text "In contexts where other standards . . .").

```
6691
```

6692 The grouping specification is based on, but not identical to, the C Standard . The "-1"  
6693 signals that no further grouping shall be performed, the equivalent of (CHAR\_MAX) in  
6694 the C Standard ).

```
6695
```

6696 The FDCC-set definition is an extension of the C Standard localeconv() specification. In  
6697 particular, rules on how currency\_symbol is treated are extended to also cover int\_  
6698 curr\_symbol, and p\_sep\_by\_space and n\_sep\_by\_space have been augmented with the  
6699 value 2, which places a space between the sign and the symbol (if they are adjacent;  
6700 otherwise it should be treated as a 0). The following table shows the result of various  
6701 combinations:

```
6702
```

```
6703
```

		p_sep_by_space			
		2	1	0	
6704					
6705					
6706					
6707	p_cs_precedes = 1	p_sign_posn = 0	(\$ 1.25)	(\$ 1.25)	(\$1.25)
6708		p_sign_posn = 1	+ \$1.25	+\$ 1.25	+\$1.25
6709		p_sign_posn = 2	\$1.25 +	\$ 1.25+	\$1.25+
6710		p_sign_posn = 3	+ \$1.25	+\$ 1.25	+\$1.25
6711		p_sign_posn = 4	\$ +1.25	+\$ 1.25	+\$1.25
6712					
6713	p_cs_precedes = 0	p_sign_posn = 0	(1.25 \$)	(1.25 \$)	(1.25\$)
6714		p_sign_posn = 1	+1.25 \$	+1.25 \$	+1.25\$
6715		p_sign_posn = 2	1.25\$ +	1.25 \$+	1.25\$+
6716		p_sign_posn = 3	1.25+ \$	1.25 +\$	1.25+\$
6717		p_sign_posn = 4	1.25\$ +	1.25 \$+	1.25\$+

```
6718
```

```
6719
```

6720 The following is an example of the interpretation of the mon\_grouping keyword.  
6721 Assuming that the value to be formatted is 123456789 and the mon\_thousands\_sep is "",  
6722 then the following table shows the result. The third column shows the equivalent C

6723 Standard string that would be used to accommodate this grouping. It is the responsibility  
 6724 of the utility to perform mappings of the formats in this clause to those used by language  
 6725 bindings such as the C Standard .  
 6726

6727	Mon_grouping	Formatted Value	C String
6728	3;-1	123456'789	"\3\177"
6729	3	123'456'789	"\3"
6730	3;2;-1	1234'56'789	"\3\2\177"
6731	3;2	12'34'56'789	"\3\2"
6732	-1	123456789	"177"

6733

6734 In these examples, the octal value of (CHAR\_MAX) is 177.  
 6735

6736

6737 The multiple currency support is specified such that a FDCC-set can be used without  
 6738 change during the transition period in a static environment. For example in the case of the  
 6739 Euro currency as being employed in a number of European countries, there is no need to  
 6740 change the FDCC-set when shifting from one currency to two concurrent currencies; and  
 6741 there is no need to change FDCC-set, when changing to the Euro as the only currency.  
 6742 Also the same application call can be made to be valid for countries with a single  
 6743 currency and countries with dual currencies. The specifications can also be used without  
 6744 change of the FDCC-set on an installation, when converting from one national currency to  
 6745 another, for example when removing some zeroes to form a new currency.  
 6746

6747

6748 The following example illustrates the support for multiple currencies; the example is for  
 the Euro in Germany:

6749

```

6750 LC_MONETARY
6751 valid_from ; 19990101
6752 valid_to 20020630;
6753 conversion_rate 1; 195/100
6754 int_curr_symbol "<D><E><M><SP>"; "<E><U><R><SP>"
6755 currency_symbol "<D><M>"; "<E><U><R>"
6756 mon_decimal_point "<,>"
6757 mon_thousands_sep "<.>"
6758 mon_grouping 3;3
6759 positive_sign ""
6760 negative_sign "<->"
6761 int_frac_digits 2; 2
6762 frac_digits 2; 2
6763 p_cs_precedes 1; 1
6764 p_sep_by_space 2; 2
6765 n_cs_precedes 1; 1
6766 n_sep_by_space 2; 2
6767 p_sign_posn 4; 4
6768 n_sign_posn 4; 4
6769
6770 END LC_MONETARY
  
```

6771

6772

### 6773 B.1.5 LC\_NUMERIC Rationale.

6774

6775 See the rationale for LC\_MONETARY (B1.3) for a description of the behaviour of  
 6776 grouping.

6777

### 6778 B.1.6 LC\_TIME Rationale.

6779

6779 The LC\_TIME descriptions of abday, day, and abmon imply a Gregorian style calendar

6780 (7-day weeks, 12-month years, leap years, etc.). Other calendars can be supported, for  
6781 example calendars with a fixed week length.

6782  
6783 In some FDCC-sets the field descriptors for weekday and month names will be given with  
6784 an initial small letter. Programs using these fields may need to adjust the capitalization if  
6785 the output is going to be used at the beginning of a sentence.

6786  
6787 The field descriptors corresponding to the optional keywords consist of a modifier  
6788 followed by a traditional field descriptor (for instance %Ex). If the optional keywords are  
6789 not supported by the application or are unspecified for the current FDCC-set, these field  
6790 descriptors shall be treated as the traditional field descriptor. For instance, assume the  
6791 following keywords:

```
6792
6793     alt_digits "0th";"1st";"2nd";"3rd";"4th";"5th";"6th";"7th";"8th";"9th";"10th"
6794     d_fmt "The %Od day of %B in %Y"
```

6795  
6796 On 7/4/1776, the %x field descriptor would result in "The 4th day of July in 1776," while  
6797 7/14/1789 would come out as "The 14 day of July in 1789." It can be noted that the above  
6798 example is for illustrative purposes only; the %o modifier is primarily intended to provide  
6799 for Kanji or Hindi digits in date formats. While it is clear that an alternate year format is  
6800 required, there is no consensus on the format or the requirements. As a result, while these  
6801 keywords are reserved, the details are left unspecified. It is expected that National  
6802 Standards Bodies will provide specifications.

6803  
6804

### 6805 **B.1.7 LC\_MESSAGES Rationale.**

6806  
6807 The LC\_MESSAGES category is described in clause 4 as affecting the language used by  
6808 utilities for their output. The mechanism used by the application to accomplish this, other  
6809 than the responses shown here in the FDCC-set definition, is not specified by this version  
6810 of this standard. The internationalization working group is developing an interface that  
6811 would allow applications (and, presumably some of the standard utilities) to access  
6812 messages from various message catalogs, tailored to a user's LC\_MESSAGES value.

6813  
6814

### 6815 **B.1.8 LC\_PAPER Rationale.**

6816  
6817 The LC\_PAPER category gives information to prepare output on a printer. Only the  
6818 physical measurement s of the height and width is available, as this is the information  
6819 most often available in various document handling applications.

6820  
6821

### 6822 **B.1.9 LC\_NAME Rationale.**

6823  
6824 The LC\_NAME category gives information to prepare a text for addressing a person, for  
6825 example as a part of a postal address on an envelope, or as a saluting line in a letter.  
6826 The information is intended to be given to an API that has the various naming information  
6827 as parameters and yields a formatted string as the return value.

6828  
6829

6830 **B.1.10 LC\_ADDRESS Rationale.**

6831  
6832 The LC\_ADDRESS category gives information to prepare a text for writing an address,  
6833 for example as a part of a postal address on an envelope. The information is intended to  
6834 be given to an API that has the various address information as parameters and yields a  
6835 formatted string as the return value.

6836  
6837

6838 **B.1.11 LC\_TELEPHONE Rationale.**

6839  
6840 The LC\_TELEPHONE category gives information to prepare a text for writing a telephone  
6841 number. The information is intended to be given to an API that has the various  
6842 information on a telephone number as parameters and yields a formatted string as the  
6843 return value. Both an international and a domestic formatting possibility is available.

6844

6845

6846 **B.2 Character Set Rationale.**

6847

6848 This standard poses no requirement that multiple character sets or code sets be supported,  
6849 leaving this as a marketing differentiation for implementors. Although multiple charmaps  
6850 are supported, it is the responsibility of the application to provide the file(s); if only one is  
6851 provided, only that one will be accessible.

6852

6853 The character set description text provides the capability to describe character set attributes  
6854 (such as collation order or character classes) independent of character set encoding, and  
6855 using only the characters in the portable character set. This makes it possible to create  
6856 "generic" FDCC-set source texts for all code sets that share the portable character set  
6857 (such as the ISO/IEC 8859 family or IBM Extended ASCII).

6858

6859 Applications are free to describe more than one code set in a character set description text.  
6860 For example, if an application defines ISO/IEC 8859-1 as the primary code set, and  
6861 ISO/IEC 8859-2 as an alternate set, with each character from the alternate code set  
6862 preceded in data by a shift code, a character set description text could contain a complete  
6863 description of the primary set and those characters from the secondary that are not  
6864 identical, the encoding of the latter including the shift code.

6865

6866 Applications are free to choose their own symbolic names, as long as the names identified  
6867 by this standard are also defined; this provides support for already existing "character  
6868 names".

6869

6870 The charmap was introduced to resolve problems with the portability of, especially,  
6871 FDCC-set sources. While the portable character set (in Table 1) is a constant across all  
6872 FDCC-sets for a particular application, this is not true for the extended character set.  
6873 However, the particular coded character set used for an application does not necessarily  
6874 imply different characteristics or collation: on the contrary, these attributes should in many  
6875 cases be identical, regardless of codeset. The charmap provides the capability to define a  
6876 common FDCC-set definition for multiple codesets (the same FDCC-set source can be  
6877 used for codesets with different extended characters; the ability in the charmap to define  
6878 "empty" names allows for characters missing in certain codesets).

6879

6880 In addition, some implementors have expressed an interest in using the charmap to define  
6881 certain other characteristics of codesets, such as the <mb\_cur\_max> value for the  
6882 particular codeset. (Note that <mb\_cur\_max> has to be equal to or lower than the C  
6883 Standard {MB\_LEN\_MAX}, which is the application limit). Such extensions are not  
6884 described here; but may be added in a later revision of this standard.

6885  
6886 The <escape\_char> declaration was added at the request of the international community to  
6887 ease the creation of portable charmaps on terminals not implementing the default  
6888 backslash escape. (This approach was adopted because this is a new interface invented by  
6889 POSIX-2. Historical interfaces, such as the shell command language and awk, have not  
6890 been modified to accommodate this type of terminal.)

6891  
6892 The octal number notation was selected to match those of POSIX "awk" and "tr" utilities  
6893 and is consistent with that used by the POSIX localedef utility.

6894  
6895 The charmap capability implements a facility available at some X/Open compatible  
6896 applications. Its prime virtue is to support "generic" collation sequence source definitions.  
6897 An implementor or an applications developer can produce a template definition that can be  
6898 used to produce several codeset-dependent "compiled" FDCC-set definitions. The facility  
6899 also removes any dependency in many source definitions on characters outside the  
6900 character set defined in this clause.

6901  
6902 The charmap allows specification of more than one encoding of a character. This allows  
6903 for encodings that can encode items in more than one way. For example, an item can be  
6904 encoded once as a fully composed character and again as a base character plus combining  
6905 character. This would allow either representation to be recognized. As only the first  
6906 occurrence of the character may be output, this technique could be used to normalize a  
6907 character stream.

6908  
6909 The ISO 2022 support introduced gives the possibility to refer other definitions via  
6910 charmaps, so the full encoding does not have to be replicated. It supports shifting with G0,  
6911 G1, G2 and G3 sets, and also general shifting of coded character sets via escape  
6912 sequences.

### 6913 6914 6915 **B.3 Repertoiremap Rationale.**

6916  
6917 The repertoiremap was introduced to make FDCC-sets independent of the availability of  
6918 charmaps. With the repertoiremap it is possible to use a FDCC-set encoded with one set of  
6919 symbolic character names, together with charmaps with other symbolic character naming  
6920 schemes, provided there are repertoiremaps available for both naming schemes.

6921  
6922 Repertoiremaps are also useful to describe repertoires of characters, to be used for  
6923 example for transliteration.

## Annex C (informative)

### BNF Grammar

#### C.1 BNF Syntax Rules

The syntax used here is near to ISO/IEC 14977, but "\_" is allowed in identifiers, and comma is not used as concatenator, as the items are just concatenated.

Definitions between (\* \*) make use of terms not defined in this BNF syntax, and assume general English usage.

Other conventions:

\* means 0 or more repetitions of a token.

Brackets [ ] indicate optional occurrence of a token.

(\* \*) are ISO/IEC 14977 comment symbols.

There may be more specifications in the normative text that describes restrictions on the grammar.

#### C.2 Grammar for FDCC-sets

```
(* The following grammar rules are common to all categories *)
CHAR = (* any character *);
graphic_char = CHAR - (* control_characters *) - space ;
space = ' ' | <TAB> ;
EOL = (* anything that makes an End Of Line (EOL)
in the operating system employed *)
| comment EOL ;
COMMENT_CHAR = (* defined by the 'comment_char' keyword *) ;
ESCAPE_CHAR = (* defined by the 'escape_char' keyword *) ;
CHARSYMBOL = simple_symbol | UCS_symbol ;
COLLSYMBOL = simple_symbol ;
COLLELEMENT = simple_symbol ;
SECTIONSYMBOL = simple_symbol ;
octdigit = '0' | '1' | '2' | '3' | '4' | '5' | '6' | '7' ;
digit = '0' | '1' | '2' | '3' | '4' | '5' | '6' | '7' | '8' | '9' ;
hex_upper = 'A' | 'B' | 'C' | 'D' | 'E' | 'F' | digit ;
hexdigit = hex_upper | 'a' | 'b' | 'c' | 'd' | 'e' | 'f' ;
letter = 'a' | 'b' | 'c' | 'd' | 'e' | 'f' | 'g' | 'h' | 'i' | 'j' | 'k'
| 'l' | 'm' | 'n' | 'o' | 'p' | 'q' | 'r' | 's'
| 't' | 'u' | 'v' | 'w' | 'x' | 'y' | 'z' | 'A' | 'B' | 'C' | 'D' | 'E'
| 'F' | 'G' | 'H' | 'I' | 'J' | 'K' | 'L' | 'M' | 'N' | 'O' | 'P' |
| 'Q' | 'R' | 'S' | 'T' | 'U' | 'V' | 'W' | 'X' | 'Y' | 'Z' ;
portable_graph = letter | digit | '!' | '"' | '#' | '$' | '%' | '&'
| "'" | '(' | ')' | '*' | '+' | ',' | '-' | '.' | '/' | ':' | ';'
| '<' | '=' | '>' | '?' | '@' | '[' | '\' | ']' | '^' | '_'
| '`' | '{' | '|' | '}' | '~' ;
portable_char = portable_graph | ' ' | <NUL> | <ALERT>
| <BACKSPACE> | <TAB> | <CARRIAGE_RETURN>
| <NEWLINE> | <VERTICAL_TAB> | <FORM_FEED> ;
OCTAL_CHAR = ESCAPE_CHAR octdigit octdigit octdigit* ;
HEX_CHAR = ESCAPE_CHAR 'x' hexdigit hexdigit hexdigit* ;
DECIMAL_CHAR = ESCAPE_CHAR 'd' digit digit digit* ;
NUMBER = digit digit* ;
id_part = letter | digit | '-' | '_' ;
four_digit_hex_string = hex_upper hex_upper hex_upper hex_upper ;
identifier = letter id_part* ;
```



```

6984 simple_symbol = space* '<' graphic_char graphic_char* '>' ;
6985 ucs_symbol = space* '<U' four_digit_hex_string
6986 [ four_digit_hex_string ] '>' ;
6987 quoted_string = ''' char_symbol* ''' ;
6988 quoted_nonempty_string = ''' char_symbol [ char_symbol* ] ''' ;
6989 char_symbol = CHAR | CHARSYMBOL
6990 | OCTAL_CHAR | HEX_CHAR | DECIMAL_CHAR ;
6991 elem_list = elem elem* ;
6992 elem = char_symbol | COLLSYMBOL | COLLELEMENT ;
6993 symb_list = COLLSYMBOL COLLSYMBOL* ;
6994 FDCC_set_name = FDCC_NAME | ''' FDCC_NAME ''' ;
6995 copy_FDCC_set = 'copy' FDCC_set_name EOL ;
6996 FDCC_NAME = char_symbol char_symbol* ;
6997 semicolon = ';' ;
6998 comment = COMMENT_CHAR CHAR* ;
6999
7000 (* The following is the overall FDCC-set grammar *)
7001 FDCC_set_definition = [ global_statement* ] category* ;
7002 global_statement = 'escape_char' character EOL
7003 | 'comment_char' character EOL
7004 | 'repertoiremap' quoted_string EOL
7005 | 'charmap' quoted_string EOL ;
7006 category = lc_identification | lc_ctype | lc_collate
7007 | lc_monetary | lc_numeric | lc_time
7008 | lc_messages | lc_paper | lc_telephone
7009 | lc_name | lc_address ;
7010
7011 (* The following is the LC_IDENTIFICATION category grammar *)
7012 lc_ident = ident_head ident_keyword* ident_tail
7013 | ident_head copy_FDCC_set ident_tail ;
7014 ident_head = 'LC_IDENTIFICATION' EOL ;
7015 ident_keyword = ident_keyword_string quoted_string EOL ;
7016 ident_keyword_string = 'title' | 'source' | 'address' | 'contact'
7017 | 'email' | 'tel' | 'fax' | 'language'
7018 | 'territory' | 'audience' | 'application'
7019 | 'abbreviation' | 'revision' | 'date' ;
7020 ident_tail = 'END' 'LC_IDENTIFICATION' EOL ;
7021
7022
7023 (* The following is the LC_CTYPE category grammar *)
7024 lc_ctype = ctype_head ctype_keyword* [ translit ]
7025 ctype_tail
7026 | ctype_head copy_FDCC_set ctype_tail ;
7027 ctype_head = 'LC_CTYPE' EOL ;
7028 ctype_keyword = charclass_keyword charclass_list EOL
7029 | charconv_keyword charconv_list EOL ;
7030 charclass_keyword = 'upper' | 'lower' | 'alpha' | 'digit'
7031 | 'punct' | 'xdigit' | 'space' | 'print'
7032 | 'graph' | 'blank' | 'cntrl' | 'outdigit'
7033 | 'class' class_name semicolon ;
7034 class_name = '"combining"' | '"combining_level3"'
7035 | ''' identifier ''' ;
7036 charclass_list = charclass_list semicolon char_symbol
7037 | charclass_list semicolon abs_ellipsis
7038 semicolon char_symbol
7039 | charclass_list semicolon CHARSYMBOL
7040 ctype_symbolic_ellipses CHARSYMBOL
7041 | char_symbol ;
7042 charconv_keyword = 'toupper' | 'tolower'
7043 | 'map' ''' identifier ''' semicolon ;
7044 charconv_list = charconv_list semicolon charconv_entry
7045 | charconv_entry ;
7046 charconv_entry = '(' char_symbol ',' char_symbol ')' ;
7047 ctype_symbolic_ellipses = '...' | '.....' | '..(2)..';
7048 ctype_abs_ellipses = '...' ;
7049 translit = translit_start [translit_include]
7050 [default_missing] translit_statement*
7051 translit_end ;
7052 translit_start = 'translit_start' EOL ;
7053 translit_include = 'include' FDCC_set_name semicolon
7054 quoted_nonempty_string EOL ;

```

```

7055 default_missing = 'default_missing' quoted_string EOL ;
7056 translit_statement = char_or_string char_or_string [ semicolon
7057 char_or_string ] EOL ;
7058 translit_end = 'translit_end' EOL ;
7059 ctype_tail = 'END' 'LC_TYPE' EOL ;
7060
7061 (* The following is the LC_COLLATE category grammar *)
7062 lc_collate = collate_head collate_keywords collate_tail ;
7063 collate_head = 'LC_COLLATE' EOL ;
7064 collate_keywords = [ opt_statement* ] order_statements ;
7065 opt_statement = 'collating-symbol' COLLSYMBOL* EOL
7066 | 'collating-element' COLLELEMENT
7067 collelem_string EOL
7068 | 'section-symbol' SECTIONSYMBOL EOL
7069 | 'copy' FDCC_set_name EOL
7070 | 'col_weight_max' NUMBER EOL
7071 | 'symbol-equivalence' COLLSYMBOL COLLSYMBOL ;
7072 collelem_string = ''' char_symbol char_symbol char_symbol* ''' ;
7073 order_statements = order_start collation_order order_end ;
7074 order_start = 'order_start' COLLSYMBOL [ semicolon
7075 order_opts ] EOL
7076 | 'order_start' [ order_opts ] EOL ;
7077 order_opts = order_opt [ semicolon order_opt ] ;
7078 order_opt = order_opt [ ',' opt_word ] ;
7079 opt_word = 'forward' | 'backward' | 'position' ;
7080 collation_order = collation_statement* ;
7081 collation_statement = COLLSYMBOL EOL
7082 | collating_element [ weight_list ] EOL ;
7083 collation_element = char_symbol | COLLELEMENT
7084 | ellipses | 'UNDEFINED' ;
7085 weight_list = weight_symbol [ semicolon weight_symbol ]* ;
7086 weight_symbol = (* empty *)
7087 | char_symbol
7088 | COLLSYMBOL
7089 | ''' elem_list '''
7090 | ''' symb_list ''' | 'IGNORE' ;
7091 ellipses = '...' | '...' | '.....' ;
7092 reorder_after = 'reorder-after' COLLSYMBOL EOL ;
7093 reorder_end = 'reorder-end' EOL ;
7094 reorder_section_after = 'reorder-section-after' SECTIONSYMBOL
7095 SECTIONSYMBOL EOL ;
7096 reorder_section_end = 'reorder-section-end' EOL ;
7097 order_end = 'order_end' EOL
7098 collate_tail = 'END' 'LC_COLLATE' EOL ;
7099
7100 (* The following is the LC_MESSAGES category grammar *)
7101 lc_messages = messages_head messages_keyword* messages_tail
7102
7103 | messages_head copy_FDCC_set messages_tail ;
7104 messages_head = 'LC_MESSAGES' EOL ;
7105 messages_keyword = 'yesexpr' ''' EXTENDED_REG_EXPR ''' EOL
7106 | 'yesexpr' ''' EXTENDED_REG_EXPR ''' EOL ;
7107 messages_tail = 'END' 'LC_MESSAGES' EOL ;
7108
7109 (* The following is the LC_MONETARY category grammar *)
7110 lc_monetary = monetary_head monetary_keyword* monetary_tail
7111 | monetary_head copy_FDCC_set monetary_tail ;
7112 monetary_head = 'LC_MONETARY' EOL ;
7113 monetary_keyword = mon_keyword_string quoted_string EOL
7114 | mon_keyword_strings mon_string_list EOL
7115 | mon_keyword_char mon_number_list EOL
7116 | mon_keyword_date mon_date_list EOL
7117 | 'conversion_rate' mon_conv_list EOL
7118 | 'mon_grouping' mon_group_list EOL ;
7119 mon_keyword_string = 'mon_decimal_point' | 'mon_thousands_sep'
7120 | 'positive_sign' | 'negative_sign' ;
7121 mon_keyword_strings = 'int_curr_symbol' | 'currency_symbol' ;
7122 mon_keyword_char = 'int_frac_digits' | 'frac_digits'
7123 | 'p_cs_precedes' | 'p_sep_by_space'

```

```

7124 | 'n_cs_precedes' | 'n_sep_by_space'
7125 | 'int_p_cs_precedes' | 'int_p_sep_by_space'
7126 | 'int_n_cs_precedes' | 'int_n_sep_by_space'
7127 | 'p_sign_posn' | 'n_sign_posn'
7128 | 'int_p_sign_posn' | 'int_n_sign_posn' ;
7129 mon_keyword_date = 'valid_from' | 'valid_to' ;
7130 mon_date_list = mon_date | mon_date_list ';' mon_date ;
7131 mon_date = [ '-' ] 8 * digit ;
7132 mon_group_list = NUMBER | mon_group_list ';' NUMBER ;
7133 mon_string_list = quoted_string [ ';' quoted_string ] * ;
7134 mon_number_list = mon_number | mon_number_list ';' mon_number ;
7135 mon_number = NUMBER | -1 ;
7136 mon_conv_list = mon_pair | mon_conv_list ';' mon_pair ;
7137 mon_pair = NUMBER '/' NUMBER ;
7138 monetary_tail = 'END' 'LC_MONETARY' EOL ;
7139
7140 (* The following is the LC_NUMERIC category grammar *)
7141 lc_numeric = numeric_head numeric_keyword* numeric_tail
7142 | numeric_head copy_FDCC_set numeric_tail ;
7143 numeric_head = 'LC_NUMERIC' EOL ;
7144 numeric_keyword = num_keyword_string quoted_string EOL
7145 | num_keyword_grouping num_group_list EOL ;
7146 num_keyword_string = 'decimal_point' | 'thousands_sep' ;
7147 num_keyword_grouping = 'grouping' ;
7148 num_group_list = NUMBER
7149 | num_group_list semicolon NUMBER ;
7150 numeric_tail = 'END' 'LC_NUMERIC' EOL ;
7151
7152 (* The following is the LC_TIME category grammar *)
7153 lc_time = time_head time_keyword* time_tail
7154 | time_head copy_FDCC_set time_tail ;
7155 time_head = 'LC_TIME' EOL ;
7156 time_keyword = time_keyword_name time_list EOL
7157 | time_keyword_fmt quoted_string EOL
7158 | time_keyword_opt time_list EOL
7159 | 'week' NUMBER semicolon mon_date semicolon
7160 NUMBER EOL
7161 | time_keyword_num NUMBER EOL
7162 | 'timezone' time_list EOL ;
7163 time_keyword_name = 'abday' | 'day' | 'abmon' | 'mon' | 'am_pm' ;
7164 time_keyword_fmt = 'd_t_fmt' | 'd_fmt' | 't_fmt' | 't_fmt_ampm' ;
7165 time_keyword_opt = 'era' | 'era_year' | 'era_d_fmt' | 'alt_digits'
7166 ;
7167 time_keyword_week = 'week' ;
7168 time_keyword_num = 'first_weekday' | 'first_workday'
7169 | 'cal_direction' ;
7170 time_list = time_list semicolon quoted_string
7171 | quoted_string ;
7172 time_tail = 'END' 'LC_TIME' EOL ;
7173
7174 (* The following is the LC_PAPER category grammar *)
7175 lc_paper = paper_head paper_keyword* paper_tail
7176 | paper_head copy_FDCC_set paper_tail ;
7177 paper_head = 'LC_PAPER' EOL ;
7178 paper_keyword = paper_keyword_num NUMBER EOL ;
7179 paper_keyword_num = 'height' | 'width' ;
7180 paper_tail = 'END' 'LC_PAPER' EOL ;
7181
7182 (* The following is the LC_NAME category grammar *)
7183 lc_name = name_head name_keyword* name_tail
7184 | name_head copy_FDCC_set name_tail ;
7185 name_head = 'LC_NAME' EOL ;
7186 name_keyword = name_keyword_string quoted_string EOL ;
7187 name_keyword_string = 'name_fmt' | 'name_gen' | 'name_mr'
7188 | 'name_mrs' | 'name_ms' | 'name_miss'
7189 | 'name_ms' ;
7190 name_tail = 'END' 'LC_NAME' EOL ;
7191
7192 (* The following is the LC_ADDRESS category grammar *)
7193 lc_address = address_head address_keyword* address_tail
7194 | address_head copy_FDCC_set address_tail ;

```

```
7195 address_head = 'LC_ADDRESS' EOL ;
7196 address_keyword = address_keyword_string quoted_string EOL
7197 | address_keyword_num NUMBER EOL ;
7198 address_keyword_string = 'postal_fmt' | 'country_name' |
7199 'country_post' | 'country_ab2' | 'country_ab3'
7200 | 'country_car' | 'country_isbn' | 'lang_name' |
7201 'lang_ab' | 'lang_term' | 'lang_lib' ;
7202 address_keyword_num = "country_num" ;
7203 address_tail = 'END' 'LC_ADDRESS' EOL ;
7204
7205 (* The following is the LC_TELEPHONE category grammar *)
7206 lc_tel = tel_head tel_keyword* tel_tail
7207 | tel_head copy_FDCC_set tel_tail ;
7208 tel_head = 'LC_TELEPHONE' EOL ;
7209 tel_keyword = tel_keyword_string quoted_string EOL ;
7210 tel_keyword_string = 'tel_int_fmt' | 'tel_dom_fmt' | 'int_select'
7211 | 'int_prefix' ;
7212 tel_tail = 'END' 'LC_TELEPHONE' EOL ;
7213
```

		<b>Annex D</b>	
		(informative)	
		<b>Index</b>	
7214			
7215			
7216			
7217			
7218			
7219	abbreviation	4.1	cntrl 4.2.1
7220	abday	4.6	code_set_name 5.1
7221	abmon	4.6	coded character 3.1.3
7222	absolute ellipses	4.2	col_weight_max 4.3, 4.3.3
7223	address	4.1	collating-element 4.3
7224	addresses	4.10	collating statements 4.3.1
7225	addset	5.1	collating-symbol 4.3.6
7226	affirmative response	3.1.17	collating element 3.1.13
7227	alpha	4.2.1	collating sequence 3.1.15
7228	alt_digits	4.6	collating-element 4.3.5
7229	am_pm	4.6	collating-symbol 4.3
7230	application	4.1	collation 3.1.12
7231	audience	4.1	combining 4.2.1
7232	blank	4.2.1	combining_level3 4.2.1
7233	block_separator	4.2.1	comment_char 4.0.2.1, 5.1
7234	byte	3.1.1	conformance 7
7235	cal_direction	4.6	contact 4.1
7236	category	4.1	control characters 4.2.1
7237	category names	4.0	conversion_rate 4.4
7238	category trailer	4.0	copy 4.1, 4.2.1, 4.3.2, 4.4, 4.5, 4.6, 4.7, 4.8, 4.9, 4.10, 4.11
7239	category header	4.0	
7240	category body	4.0	country_ab2 4.10
7241	character	3.1.2	country_ab3 4.10
7242	character, graphic	4.2.1	country_car 4.10
7243	character, special	4.2.1	country_isbn 4.10
7244	character representation	4.0.1	country_name 4.10
7245	character, native digit	4.2.1	country_num 4.10
7246	character, hexadecimal digit	4.2.1	country_post 4.10
7247	character, multibyte	4.0.1	cultural convention 3.1.5
7248	character, decimal constant	4.0.1	currency_symbol 4.4
7249	character, hexadecimal constant	4.0.1	d_fmt 4.6
7250	character, space	4.2.1	d_t_fmt 4.6
7251	character, octal constant	4.0.1	date field descriptors 4.6.1
7252	character, control	4.2.1	date 4.1
7253	character, blank	4.2.1	day 4.6
7254	character, digit	4.2.1	decimal_point 4.5
7255	character, punctuation	4.2.1	default_missing 4.2.2
7256	character, printable	3.1.10	define 4.3.14.0, 4.3
7257	character class	3.1.9	definitions 3.1
7258	character, coded	3.1.3	digit 4.2.1
7259	Character set rationale	B.2	elif 4.3.14.6, 4.3
7260	charmap text	5.1	ellipses 4.2, 4.3.1, 5.1
7261	charmap	5, 4.0.2.4, 3.1.7	ellipses, absolute 4.2, 4.3.1
7262	charmap rationale	B.2	ellipses, symbolic 4.2, 4.3.1, 5.1
7263	class	4.2.1	else 4.3, 4.3.14.5

7264	email	4.1	LC_CTYPE	4.2
7265	endif	4.3	LC_CTYPE rationale	B.1.2
7266	endif	4.3.14.7	LC_IDENTIFICATION	4.1
7267	equivalence class	3.1.16	LC_IDENTIFICATION rationale	B.1.1
7268	era	4.6	LC_MESSAGES	4.7
7269	era_d_fmt	4.6	LC_MESSAGES rationale	B.1.7
7270	era_year	4.6	LC_MONETARY	4.4
7271	escape_char	4.0.2.2, 5.1, 6	LC_MONETARY rationale	B.1.4
7272	esqseq	5.1	LC_NAME	4.9
7273	euro	B.1.3	LC_NAME rationale	B.1.9
7274	extended regular expression	4.7	LC_NUMERIC	4.5
7275	fax	4.1	LC_NUMERIC rationale	B.1.5
7276	FDCC-set, definition	4.0	LC_PAPER	4.8
7277	FDCC-set	4f	LC_PAPER rationale	B.1.8
7278	FDCC-set	3.1.6	LC_TELEPHONE	4.11
7279	FDCC-set rationale	B.1	LC_TELEPHONE rationale	B.1.11
7280	first_weekday	4.6	LC_TIME	4.6
7281	first_workday	4.6	LC_TIME rationale	B.1.6
7282	frac_digits	4.4	LC_X_	4
7283	graph	4.2.1	line continuation	3.2.2
7284	graphic chracters	4.2.1	lower	4.2.1
7285	grouping	4.5	map	4.2.1
7286	height	4.8	mb_cur_max	5.1
7287	ifdef	4.3.14.3	mb_cur_min	5.1
7288	ifdef	4.3	messages	4.7
7289	ifndef	4.3	modified date field descriptors	4.6.2
7290	ifndef	4.3.14.4	mon	4.6
7291	include	4.2.2	mon_decimal_point	4.4
7292	include	5.1	mon_grouping	4.4
7293	include	4.2.2.2	mon_thousands_sep	4.4
7294	int_curr_symbol	4.4	monetary	4.4
7295	int_frac_digits	4.4	multicharacter collating element	3.1.14
7296	int_n_cs_precedes	4.4	n_cs_precedes	4.4
7297	int_n_sep_by_space	4.4	n_sep_by_space	4.4
7298	int_n_sign_posn	4.4	n_sign_posn	4.4
7299	int_p_cs_precedes	4.4	name formatting	4.9
7300	int_p_sep_by_space	4.4	name_fmt	4.9
7301	int_p_sign_posn	4.4	name_gen	4.9
7302	int_prefix	4.11	name_miss	4.9
7303	int_select	4.11	name_mr	4.9
7304	keywords	4.0	name_mrs	4.9
7305	lang_ab	4.10	name_ms	4.9
7306	lang_lib	4.10	negative response	3.1.18
7307	lang_name	4.10	negative_sign	4.4
7308	lang_term	4.10	noexpr	4.7
7309	language	4.1	normal_connect	4.2.1
7310	LC_ADDRESS	4.10	notations	3.2
7311	LC_ADDRESS rationale	B.1.10	numeric	4.5
7312	LC_COLLATE	4.3	operands	4.0
7313	LC_COLLATE rationale	B.1.3	order_end	4.3.9, 4.3

7314	order_start	4.3, 4.3.8	timezone	4.6
7315	outdigit	4.2.1	title	4.1
7316	p_cs_precedes	4.4	toggling keywords	4.3.14
7317	p_sep_by_space	4.4	tolower	4.2.1
7318	p_sign_posn	4.4	tosymmetric	4.2.1
7319	paper format	4.8	toupper	4.2.1
7320	portable character set	3.2.4	translit_end	4.2.2
7321	positive_sign	4.4	translit_start	4.2.2
7322	POSIX	1	transliteration	4.2.2
7323	POSIX differences	A	transliteration statements	4.2.2.1
7324	POSIX conformance	4.1	undef	4.3, 4.3.14.2
7325	postal addresses	4.10	upper	4.2.1
7326	postal_fmt	4.10	valid_from	4.4
7327	pre-category statements	4.0.2	valid_to	4.4
7328	print	4.2.1	visible glyph portable characters	3.2.4
7329	printable character	3.1.10	week	4.6
7330	punct	4.2.1	white space	3.1.11
7331	punctuation characters	4.2.1	width	4.8
7332	references	2	xdigit	4.2.1
7333	reorder-section-end	4.3.13	yesexpr	4.7
7334	reorder-section-after	4.3.12		
7335	reorder-section-after	4.3		
7336	reorder-after	4.3		
7337	reorder-end	4.3		
7338	reorder-section-end	4.3		
7339	reorder-after	4.3.10		
7340	reorder-end	4.3.11		
7341	reorder-after rationale	B.1.2.1		
7342	repertoire rationale	B.3		
7343	repertoire	6		
7344	repertoiremap	6, 3.1.8, 5.1, 4.0.2.3		
7345	revision	4.1		
7346	scope	1		
7347	section	4.3, 4.3.4		
7348	source	4.1		
7349	space	4.2.1		
7350	special characters	4.2.1		
7351	symbol-equivalence	4.3, 4.3.7		
7352	symbolic ellipses	4.2, 5.1		
7353	symbolic name	4.0.1		
7354	syntax format	3.2.1		
7355	t_fmt	4.6		
7356	t_fmt_ampm	4.6		
7357	tel	4.1		
7358	tel_dom_fmt	4.11		
7359	tel_int_fmt	4.11		
7360	telephone numbers	4.11		
7361	territory	4.1		
7362	text file	3.1.4		
7363	thousands_sep	4.5		

**BIBLIOGRAPHY**

- 7364  
7365  
7366 The following specifications are considered relevant to this standard, in addition to the  
7367 normative references.  
7368  
7369 CEPT, CEPT-MAILCODE, "Country code for mail"  
7370  
7371 ISO 639, "Code for the representation of names of languages"  
7372  
7373 ISO 646, "Information technology - ISO 7-bit coded character set for information inter-  
7374 change"  
7375  
7376 ISO 3166, "Code for the representation of names of countries"  
7377  
7378 ISO/IEC 9899, "Information technology - Programming Language C".  
7379  
7380 ISO/IEC 14977, "Information technology - Syntactic metalanguage - Extended BNF".  
7381  
7382 The Unicode Consortium: "The Unicode Standard, Version 2.0", Addison Wesley  
7383 Developers Press, July 1996. ISBN 0-201-48345-9.  
7384  
7385 IBM: "National Language Design Guide Volume 2 - National Language Support  
7386 Reference Manual", IBM SE09-8002-03, August 1994.  
7387  
7388 STRÍ: "Nordic Cultural Requirements on Information Technology (Summary report)",  
7389 STRÍ TS3, Libris, Reykjavík, Iceland 1992. ISBN 9979-9004-3-1.