

Reference number of working document: **ISO/IEC JTC1/SC22/WG20 N801**

Date: 2000-11-15

Reference number of document: **ISO/IEC draft2 DTR 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

H:\IPS\SAMARIN\DISKETTE\BASICEN.DOT ISO Basic template Version 3.0 1997-02-03

	Contents	Page
2		
3		
4	1 SCOPE	1
5	2 NORMATIVE REFERENCES	1
6	3 TERMS, DEFINITIONS AND NOTATIONS	2
7	4 FDCC-set	6
8	4.1 FDCC-set definition	6
9	4.2 LC_IDENTIFICATION	10
10	4.3 LC_CTYPE	11
11	4.4 LC_COLLATE	27
12	4.5 LC_MONETARY	42
13	4.6 LC_NUMERIC	46
14	4.7 LC_TIME	47
15	4.8 LC_MESSAGES	53
16	4.9 LC_XLITERATE	53
17	4.10 LC_NAME	55
18	4.11 LC_ADDRESS	57
19	4.12 LC_TELEPHONE	57
20	5 CHARMAP	58
21	6 REPERTOIREMAP	62
22	7 CONFORMANCE	89
23	Annex A (informative) DIFFERENCES FROM POSIX	90
24	Annex B (informative) RATIONALE	92
25	Annex C (informative) BNF GRAMMAR	106
26	Annex D (informative) ISSUES	106
27	Annex E (informative) INDEX	111
28	BIBLIOGRAPHY	114

## 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. In the field of information technology, ISO and IEC have established a joint  
39      technical committee, ISO/IEC JTC 1.

40  
41      The main task of a technical committee is to prepare International Standards but in  
42      exceptional circumstances, the publication of a Technical Report of one of the following  
43      types may be proposed:

- 44  
45      - type 1, when the required support cannot be obtained for the publication of an  
46      International Standard, despite repeated efforts;
- 47  
48      - type 2, when the subject is still under technical development or where for any  
49      other reason there is the future but not immediate possibility of an agreement on an  
50      International Standard;
- 51  
52      - type 3, when a technical committee has collected data of a different kind from  
53      that which is normally published as an International Standard ("state of the art", for  
54      example).

55  
56      Technical Reports are drafted in accordance with the rules given in the ISO/IEC  
57      Directives, Part 3.

58  
59      Technical Reports of types 1 and 2 are subject to review within three years of publication,  
60      to decide whether they can be transformed into International Standards. Technical Report  
61      of type 3 do not necessarily have to be reviewed until the date they provide are considered  
62      to be no longer valid or useful.

63  
64      ISO/IEC TR 14652 is a Technical Report type 1, and it was prepared by Joint Technical  
65      Committee ISO/IEC JTC 1, *Information technology, Subcommittee 22, Programming*  
66      *languages, their environments and system software interfaces.*

67  
68      The Annexes A, B, C, D and E of this Technical Report are for information only.

## 69      **Introduction**

70  
71      This Technical Report defines a general mechanism to specify cultural conventions, and it  
72      defines formats for a number of specific cultural conventions in the areas of character  
73      classification and conversion, sorting, number formatting, monetary formatting, date  
74      formatting, message display, addressing of persons, postal address formatting, and  
75      telephone number handling.

76  
77      There are a number of benefits coming from this Technical Report:

78      Rigid specification

79              Using this Technical Report, a user can rigidly specify a  
80              number of the cultural conventions that apply to the  
81              information technology environment of the user.

82      Cultural adaptability

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

89      Productivity

90              This Technical Report specifies those cultural  
91              conventions and how to specify data for them. With that  
92              data an application developer is relieved from getting the  
93              different information to support all the cultural  
94              environments for the expected customers of the product.  
95              The application developer is thus ensured of culturally  
96              correct behavior as specified by the customer, and  
97              possibly more markets may be reached as customers may  
98              have the possibility to provide the data themselves for  
99              markets that were not targeted.

100     Uniform behaviour

101              When a number of applications share one cultural  
102              specification, which may be supplied from the user or  
103              provided by the application or operating system, their  
104              behaviour for cultural adaptation becomes uniform.

105  
106     The specification format is independent of platforms and specific encoding, and targeted to  
107     be usable from a wide range of programming languages.

108  
109     A number of cultural conventions, such as spelling, hyphenation rules and terminology, are  
110     not specifiable with this Technical Report, but it provides mechanisms to define new  
111     categories and also new keywords within existing categories. An internationalized  
112     application may take advantage of information provided with the FDCC-set (such as the  
113     language) to provide further internationalized services to the user.

114  
115     This Technical Report defines a format compatible with the one used in the International  
116     string ordering standard, ISO/IEC 14651. This Technical Report is upward compatible  
117     with the ISO/IEC 9945-2:1993 POSIX shell and utilities standard, particularly its clauses  
118     2.4 and 2.5. The major extensions from that text are listed in annex A. This Technical  
119     Report has enhanced functionality in a number of areas such as ISO/IEC 10646 support,  
120     more classification of characters, transliteration, dual (multi) currency support, enhanced

date and time formatting, personal name writing, postal address formatting, telephone number handling, and management of categories. There is enhanced support for character sets including ISO/IEC 2022 handling and an enhanced method to separate the specification of cultural conventions from an actual encoding via a description of the character repertoire employed. A standard set of values for all the categories has been defined covering the repertoire of ISO/IEC 10646-1, as referenced in the normative references clause.

The Technical report was originally scheduled for adoption as an International Standard, but a number of members of ISO and IEC found the specification problematical. It was then decided to convert the specification into a Technical Report type I. Annex D lists a number of issues that some members of ISO and IEC have with the specification.

## 134 Information technology — Specification method for cultural 135 conventions

### 136 137 1 SCOPE

138 This Technical Report specifies a description format for the specification of cultural  
139 conventions, a description format for character sets, and a description format for binding  
140 character names to ISO/IEC 10646, plus a set of default values for some of these items.  
141

142 The specification is upward compatible with POSIX locale specifications - a locale  
143 conformant to POSIX specifications will also be conformant to the specifications in this  
144 Technical Report, while the reverse condition will not hold. The descriptions are intended  
145 to be coded in text files to be used via Application Programming Interfaces, that are  
146 expected to be developed for a number of programming languages.  
147

### 148 149 2 NORMATIVE REFERENCES

150 The following normative documents contain provisions which, through reference in this  
151 text, constitute provisions of this Technical Report. For dated references, subsequent  
152 amendments to, or revisions of, any of these publications do not apply. However, parties  
153 to agreements based on this Technical Report are encouraged to investigate the possibility  
154 of applying the most recent editions of the normative documents indicated below. For  
155 undated references, the latest edition of the normative document referred to applies.  
156 Members of ISO and IEC maintain registers of currently valid Technical Reports.  
157

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

159 ISO/IEC 2022, *Information technology - Character code structure and extension tech-*  
160 *niques*.

161 ISO 3166 (all parts), *Codes for the representation of names of countries and their*  
162 *subdivisions*.

163 ISO 4217, *Codes for the representation of currencies and funds*.

164 ISO 8601, *Data elements and interchange formats - Information interchange - Represen-*  
165 *tation of dates and times*.

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

168 ISO/IEC 10646-1:1993, *Information technology - Universal Multiple-Octet Coded Cha-*  
169 *racter Set (UCS) - Part 1: Architecture and Basic Multilingual Plane, including Cor.1 and*  
170 *AMD 1-9 plus AMD 18*. From AMD 18 only the characters U20AC EURO SIGN and  
171 UFFFC OBJECT REPLACEMENT CHARACTER are accounted for in this TR.

172 ISO/IEC 14651:2000, *Information technology - International string ordering - Method for*  
173 *comparing character strings and description of a default tailorable ordering*.

174 ISO/IEC 15897:1999, *Information technology - Procedures for registration of cultural*  
175 *conventions*.

### 3 TERMS, DEFINITIONS AND NOTATIONS

#### 3.1 Terms and definitions

For the purposes of this Technical Report, the terms and definitions given in the following apply.

##### 3.1.1 Bytes and characters

###### 3.1.1.1

**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 implementation defined. The least significant bit is called the low-order bit; the most significant bit is called the high-order bit.

###### 3.1.1.2

**character:**

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

###### 3.1.1.3

**coded character:**

A sequence of one or more bytes representing a single character.

###### 3.1.1.4

**text file:**

A file that contains characters organized into one or more lines.

#### 3.1.2 cultural and other major concepts

##### 3.1.2.1

**cultural convention:**

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

##### 3.1.2.2

**FDCC**

A Formal Definition of a Cultural Convention, that is a cultural convention put into a formal definition scheme.

##### 3.1.2.3

**FDCC-set:**

A Set of Formal Definitions of Cultural Conventions (FDCC's). 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.

##### 3.1.2.4

**charmap:**

A definition of a mapping between symbolic character names and character codes, plus

238 related information.  
239

240 **3.1.2.5**

241 **repertoiremap:**

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

244 **3.1.3 FDCC categories related**

245 **3.1.3.1**

246 **character class:**

247 A named set of characters sharing an attribute associated with the name of the class.

248 **3.1.3.2**

249 **collation:**

250 The logical ordering of strings according to defined precedence rules.

251 **3.1.3.3**

252 **collating element:**

253 The smallest entity used to determine logical ordering.

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

257 **3.1.3.4**

258 **multicharacter collating element:**

259 A sequence of two or more characters that collate as an entity.

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

262 **3.1.3.5**

263 **collating sequence:**

264 The relative order of collating elements as determined by the setting of the LC\_COLLATE  
265 category in the applied FDCC-set.

266 **3.1.3.6**

267 **equivalence class:**

268 A set of collating elements with the same primary collation weight.

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

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

273 **3.2 Notations**

274 The following notations and common conventions for specifications apply to this  
275 Technical Report:

276 **3.2.1 Notation for defining syntax**

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

The syntax notation looks as follows:

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

The <format> is given in a format string enclosed in double quotes, followed by a number of parameters, separated by commas. It is similar to the format specification defined in clause 2.12 in the ISO/IEC 9945-2:1993 standard and the format specification used in C language printf() function. The format of each parameter is given by an escape sequence as follows:

305        %s	specifies a string
306        %d	specifies a decimal integer
307        %c	specifies a character
308        %o	specifies an octal integer
309        %x	specifies a hexadecimal integer

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

All other characters in the format string except

316        %%	specifies a single %
317        \n	specifies an end-of-line

represent themselves.

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

### 3.2.3 Portable character set

A set of symbolic names for characters in Table 1, which is called the portable character set, is used in character description text of this specification. The first eight entries in Table 1 are defined in ISO/IEC 6429 and the rest is defined in ISO/IEC 9945-2 with some definitions from ISO/IEC 10646-1.

**Table 1: Portable character set**

Symbolic name	Glyph	UCS	Description
<NUL>		<U0000>	NULL (NUL)
<alert>		<U0007>	BELL (BEL)
<backspace>		<U0008>	BACKSPACE (BS)
<tab>		<U0009>	CHARACTER TABULATION (HT)
<carriage-return>		<U000D>	CARRIAGE RETURN (CR)
<newline>		<U000A>	LINE FEED (LF)
<vertical-tab>		<U000B>	LINE TABULATION (VT)
<form-feed>		<U000C>	FORM FEED (FF)
<space>		<U0020>	SPACE
<exclamation-mark>	!	<U0021>	EXCLAMATION MARK
<quotation-mark>	"	<U0022>	QUOTATION MARK

347	<number-sign>	#	<U0023>	NUMBER SIGN
348	<dollar-sign>	\$	<U0024>	DOLLAR SIGN
349	<percent-sign>	%	<U0025>	PERCENT SIGN
350	<ampersand>	&	<U0026>	AMPERSAND
351	<apostrophe>	,	<U0027>	APOSTROPHE
352	<left-parenthesis>	(	<U0028>	LEFT PARENTHESIS
353	<right-parenthesis>	)	<U0029>	RIGHT PARENTHESIS
354	<asterisk>	*	<U002A>	ASTERISK
355	<plus-sign>	+	<U002B>	PLUS SIGN
356	<comma>	,	<U002C>	COMMA
357	<hyphen-minus>	-	<U002D>	HYPHEN-MINUS
358	<hyphen>	-	<U002D>	HYPHEN-MINUS
359	<full-stop>	.	<U002E>	FULL STOP
360	<period>	.	<U002E>	FULL STOP
361	<slash>	/	<U002F>	SOLIDUS
362	<solidus>	/	<U002F>	SOLIDUS
363	<zero>	0	<U0030>	DIGIT ZERO
364	<one>	1	<U0031>	DIGIT ONE
365	<two>	2	<U0032>	DIGIT TWO
366	<three>	3	<U0033>	DIGIT THREE
367	<four>	4	<U0034>	DIGIT FOUR
368	<five>	5	<U0035>	DIGIT FIVE
369	<six>	6	<U0036>	DIGIT SIX
370	<seven>	7	<U0037>	DIGIT SEVEN
371	<eight>	8	<U0038>	DIGIT EIGHT
372	<nine>	9	<U0039>	DIGIT NINE
373	<colon>	:	<U003A>	COLON
374	&ltsemicolon>	;	<U003B>	SEMICOLON
375	<less-than-sign>	<	<U003C>	LESS-THAN SIGN
376	<equals-sign>	=	<U003D>	EQUALS SIGN
377	<greater-than-sign>	>	<U003E>	GREATER-THAN SIGN
378	<question-mark>	?	<U003F>	QUESTION MARK
379	<commercial-at>	@	<U0040>	COMMERCIAL AT
380	<A>	A	<U0041>	LATIN CAPITAL LETTER A
381	<B>	B	<U0042>	LATIN CAPITAL LETTER B
382	<C>	C	<U0043>	LATIN CAPITAL LETTER C
383	<D>	D	<U0044>	LATIN CAPITAL LETTER D
384	<E>	E	<U0045>	LATIN CAPITAL LETTER E
385	<F>	F	<U0046>	LATIN CAPITAL LETTER F
386	<G>	G	<U0047>	LATIN CAPITAL LETTER G
387	<H>	H	<U0048>	LATIN CAPITAL LETTER H
388	<I>	I	<U0049>	LATIN CAPITAL LETTER I
389	<J>	J	<U004A>	LATIN CAPITAL LETTER J
390	<K>	K	<U004B>	LATIN CAPITAL LETTER K
391	<L>	L	<U004C>	LATIN CAPITAL LETTER L
392	<M>	M	<U004D>	LATIN CAPITAL LETTER M
393	<N>	N	<U004E>	LATIN CAPITAL LETTER N
394	<O>	O	<U004F>	LATIN CAPITAL LETTER O
395	<P>	P	<U0050>	LATIN CAPITAL LETTER P
396	<Q>	Q	<U0051>	LATIN CAPITAL LETTER Q
397	<R>	R	<U0052>	LATIN CAPITAL LETTER R
398	<S>	S	<U0053>	LATIN CAPITAL LETTER S
399	<T>	T	<U0054>	LATIN CAPITAL LETTER T
400	<U>	U	<U0055>	LATIN CAPITAL LETTER U
401	<V>	V	<U0056>	LATIN CAPITAL LETTER V
402	<W>	W	<U0057>	LATIN CAPITAL LETTER W
403	<X>	X	<U0058>	LATIN CAPITAL LETTER X
404	<Y>	Y	<U0059>	LATIN CAPITAL LETTER Y
405	<Z>	Z	<U005A>	LATIN CAPITAL LETTER Z
406	<left-square-bracket>	[	<U005B>	LEFT SQUARE BRACKET
407	<backslash>	\	<U005C>	REVERSE SOLIDUS
408	<reverse-solidus>	\	<U005C>	REVERSE SOLIDUS
409	<right-square-bracket>	]	<U005D>	RIGHT SQUARE BRACKET
410	<circumflex-accent>	^	<U005E>	CIRCUMFLEX ACCENT
411	<circumflex>	^	<U005E>	CIRCUMFLEX ACCENT
412	<low-line>	—	<U005F>	LOW LINE
413	<underscore>	—	<U005F>	LOW LINE
414	<grave-accent>	ˋ	<U0060>	GRAVE ACCENT
415	<a>	a	<U0061>	LATIN SMALL LETTER A
416	<b>	b	<U0062>	LATIN SMALL LETTER B

417	<c>	c	<U0063>	LATIN SMALL LETTER C
418	<d>	d	<U0064>	LATIN SMALL LETTER D
419	<e>	e	<U0065>	LATIN SMALL LETTER E
420	<f>	f	<U0066>	LATIN SMALL LETTER F
421	<g>	g	<U0067>	LATIN SMALL LETTER G
422	<h>	h	<U0068>	LATIN SMALL LETTER H
423	<i>	i	<U0069>	LATIN SMALL LETTER I
424	<j>	j	<U006A>	LATIN SMALL LETTER J
425	<k>	k	<U006B>	LATIN SMALL LETTER K
426	<l>	l	<U006C>	LATIN SMALL LETTER L
427	<m>	m	<U006D>	LATIN SMALL LETTER M
428	<n>	n	<U006E>	LATIN SMALL LETTER N
429	<o>	o	<U006F>	LATIN SMALL LETTER O
430	<p>	p	<U0070>	LATIN SMALL LETTER P
431	<q>	q	<U0071>	LATIN SMALL LETTER Q
432	<r>	r	<U0072>	LATIN SMALL LETTER R
433	<s>	s	<U0073>	LATIN SMALL LETTER S
434	<t>	t	<U0074>	LATIN SMALL LETTER T
435	<u>	u	<U0075>	LATIN SMALL LETTER U
436	<v>	v	<U0076>	LATIN SMALL LETTER V
437	<w>	w	<U0077>	LATIN SMALL LETTER W
438	<x>	x	<U0078>	LATIN SMALL LETTER X
439	<y>	y	<U0079>	LATIN SMALL LETTER Y
440	<z>	z	<U007A>	LATIN SMALL LETTER Z
441	<left-brace>	{	<U007B>	LEFT CURLY BRACKET
442	<left-curly-bracket>	{	<U007B>	LEFT CURLY BRACKET
443	<vertical-line>		<U007C>	VERTICAL LINE
444	<right-brace>	}	<U007D>	RIGHT CURLY BRACKET
445	<right-curly-bracket>	}	<U007D>	RIGHT CURLY BRACKET
446	<tilde>	~	<U007E>	TILDE

This Technical Report may use other symbolic character names than the above in examples, to illustrate the use of the range of symbols allowed by the syntax specified in 4.1.1.

## 4 FDCC-set

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

460	LC_IDENTIFICATION	Versions and status of categories
461	LC_CTYPE	Character classification, case conversion and code transformation.
462	LC_COLLATE	Collation order.
463	LC_TIME	Date and time formats.
464	LC_NUMERIC	Numeric, non-monetary formatting.
465	LC_MONETARY	Monetary formatting.
466	LC_MESSAGES	Formats of informative and diagnostic messages and interactive responses.
467	LC_XLITERATE	Character transliteration.
468	LC_NAME	Format of writing personal names.
469	LC_ADDRESS	Format of postal addresses.
470	LC_TELEPHONE	Format for telephone numbers, and other telephone information.

Note: In future editions of this Technical Report further categories may be added.

477 Other category names beginning with the 3 characters "LC\_" are reserved for future  
478 standardization, except for category names beginning with the five characters "LC\_X\_"  
479 which is not used for future addition of categories specified in this Technical Report. An  
480 application may thus use category names beginning with the five characters "LC\_X\_" for  
481 application defined categories to avoid clashes with future standardized categories.  
482

483 This Technical Report also defines an FDCC-set named "i18n" with values for some of  
484 the above categories in order to simplify FDCC-set descriptions for a number of cultures.  
485 The contents of "i18n" categories should not necessarily be considered as the most  
486 commonly accepted values, while in many cases it could be the recommended values.  
487

#### 488 4.1 FDCC-set description 489

490 FDCC-sets are described with the syntax presented in this subclause. For the purposes of  
491 this Technical Report, the text is referred to as the FDCC-set definition text or FDCC-set  
492 source text.  
493

494 The **FDCC-set definition text** contains one or more FDCC-set category source definitions,  
495 and does not contain more than one definition for the same FDCC-set category. If the text  
496 contains source definitions for more than one category, application-defined categories, if  
497 present, appears after the categories defined by this clause. A category source definition  
498 contains either the definition of a category or a copy directive. In the event that some of  
499 the information for a FDCC-set category, as specified in this Technical Report, is missing  
500 from the FDCC-set source definition, the behaviour of that category, if it is referenced, is  
501 unspecified. A FDCC-set category is the normal way of specifying a single FDCC.  
502

503 There are no **naming conventions** for FDCC-sets specified in this Technical Report, but  
504 clause 6.8 in ISO/IEC 15897:1999 specifies naming rules for POSIX locales, charmaps  
505 and repertoiremaps, that may also be applied to FDCC-sets, charmaps and repertoiremaps  
506 specified according to this Technical Report.  
507

508 A **category source definition** consists of a category header, a category body, and a  
509 category trailer. A category header consists of the character string naming of the category,  
510 beginning with the characters "LC\_". The category trailer consists of the string "END",  
511 followed by one or more "blank"s and the string used in the corresponding category  
512 header.  
513

514 The **category body** consists of one or more lines of text. Each line is one of the  
515 following:  
516

- 517 - a line containing an identifier, optionally followed by one or more operands. Identifiers  
518 are either keywords, identifying a particular FDCC, or collating elements, or section  
519 symbols,
- 520 - one of transliteration statements defined in 4.3.

521 In addition to the keywords defined in this Technical Report, the source can contain  
522 application-defined keywords. Each **keyword** within a category has a unique name (i.e.,  
523 two categories can have a commonly-named keyword); no keyword starts with the  
524 characters "LC\_". Identifiers are separated from the operands by one or more "blank"s.  
525

526 **Operands** are characters, collating elements, section symbols, or strings of characters.  
527 Strings are enclosed in double-quotes. Literal double-quotes within strings are preceded by  
528

529 the <escape character>, described below. When a keyword is followed by more than one  
530 operand, the operands are separated by semicolons; "blank"s are allowed before and/or  
531 after a semicolon.

#### 534 4.1.1 Character representation

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

541 (0) The left angle bracket (<) is a reserved symbol, denoting the  
542 start of a symbolic name; when used to represent itself  
543 outside a symbolic name it is preceded by the escape  
544 character.

546 (1) A character can be represented via a **symbolic name**,  
547 enclosed within angle brackets (< and >). The symbolic  
548 name, including the angle brackets, exactly matches a  
549 symbolic name defined in a charmap or a repertoiremap to  
550 be used, and is replaced by a character value determined  
551 from the value associated with the symbolic name in the  
552 charmap or a value associated via a repertoiremap.  
553 Repertoiremaps have predefined symbolic names for UCS  
554 characters, see clause 6. A FDCC-set may also use the UCS  
555 notation of clause 6 to represent characters, without a  
556 repertoiremap being defined for the FDCC-set. Use of the  
557 escape character or a right angle bracket within a symbolic  
558 name is invalid unless the character is preceded by the  
559 escape character.

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

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

569 (2) A character can be represented by the character itself, in  
570 which case the value of the character is application-defined.  
571 Within a string, the double-quote character, the escape  
572 character, and the right angle bracket character are escaped  
573 (preceded by the escape character) to be interpreted as the  
574 character itself. Outside strings, the characters

576 , ; < > escape\_char

578 are escaped by the escape character to be interpreted as the character itself.

579 Example: c ä "May"

581                   (3)                   A character can be represented as an octal constant. An octal  
582                   constant is specified as the escape character followed by two  
583                   or more octal digits. Each constant represents a byte value.

584  
585                   Example: \143; \347; "\115"

586  
587                   (4)                   A character can be represented as a hexadecimal constant. A  
588                   hexadecimal constant is specified as the escape character  
589                   followed by an x followed by two or more hexadecimal  
590                   digits. Each constant represents a byte value.

591  
592                   Example: \x63;\xe7;

593  
594                   (5)                   A character can be represented as a decimal constant. A  
595                   decimal constant is specified as the escape character  
596                   followed by a d followed by two or more decimal digits.  
597                   Each constant represents a byte value.

598  
599                   Example: \d99; \d231;

600  
601                   (6)                   Multibyte characters can be represented by concatenated  
602                   constants specified in byte order with the last constant  
603                   specifying the least significant byte of the character.  
604                   Concatenated constants can include a mix of the above  
605                   character representations.

606  
607                   Example: \143\xe7; "\115\xe7\d171"

608  
609                   Only characters existing in the character set for which the FDCC-set definition is created  
610                   are specified, whether using symbolic names, the characters themselves, or octal, decimal,  
611                   or hexadecimal constants. If a charmap is present, only characters defined in the charmap  
612                   can be specified using octal, decimal, or hexadecimal constants. Symbolic names not  
613                   present in the charmap can be specified and are ignored, as specified under item (1)  
614                   above.

615  
616                   Note: The <character> symbolic character notation is recommended for use of specifying  
617                   all characters in a FDCC-set, to facilitate portability of the FDCC-sets, as the coded  
618                   character set of the application of the FDCC-set may be different from the coded character  
619                   set of the FDCC-set source. This is also recommended for format effectors in strings, such  
620                   as in LC\_DATE or LC\_ADDRESS, where the format effectors are allowed to be stored  
621                   together with the rest of the string, in a binary string with a different encoding from that  
622                   of the source FDCC-set.

623  
624                   **4.1.2 Continuation of lines**

625  
626                   A line in a specification can be continued by placing an escape character as the last visible  
627                   graphic character on the line; this continuation character is discarded from the input. The  
628                   line is continued to the next non-comment line.

629  
630                   **4.1.3 Names for copy keyword**

631  
632                   In most of the categories a "copy" keyword is allowed. The name specified with this copy  
633                   keyword is one of:

- 634 - "i18n" which indicate the "i18n" FDCC-set defined in this specification,  
635 - the name of a FDCC-set or POSIX locale registered by the process defined in ISO/IEC  
636 15897,  
637 - any other name which may be recognized in some local context - not being  
638 recommended as an international specification.

639  
640 **4.1.4 Pre-category statements**

641 In a FDCC-set the following statements can precede category specifications, and they  
642 apply to all categories in the specified FDCC-set.  
643

644  
645 **4.1.4.1 comment\_char**

646 The following line in a FDCC-set modifies the comment character. It has the following  
647 syntax, starting in column 1:  
648

649 "comment\_char %c\n", <comment\_character>  
650

651 The comment character defaults to the number-sign (#). All examples in this Technical  
652 Report use "%" as the <comment\_character>, except where otherwise noted. Blank lines  
653 and lines containing the <comment\_character> in the first position are ignored. In collating  
654 statements a <comment\_character> occurring where the delimiter ";" may occur,  
655 terminates the collating statement.  
656

657  
658 **4.1.4.2 escape\_char**

659 The following line in a FDCC-set modifies the escape character to be used in the text. It  
660 has the following syntax, starting in column 1:  
661

662 "escape\_char %c\n", <escape\_character>  
663

664 The escape character is used for representing characters in 4.1.1 and for continuing lines.  
665 The escape character defaults to backslash "\". All examples in this Technical Report uses  
666 "/" as the escape character, except where otherwise noted.  
667

668  
669 **4.1.4.3 repertoiremap**

670 The following line in a FDCC-set specifies the name of a repertoiremap used to define the  
671 symbolic character names in the FDCC-set. There may be at most one "repertoiremap"  
672 line. It has the following syntax, starting in column 1:  
673

674 "repertoiremap %s\n", <repertoiremap>  
675

676 The name is one of:  
677

- 678 - "i18nrep" which indicates the "i18nrep" repertoiremap defined in this specification,  
679 - the name of a <repertoiremap> registered by the process defined in ISO/IEC 15897,  
680 - any other name which may be recognized in some local context - not being  
681 recommended as an international specification.

682  
683 **4.1.4.4 charmap**

684 The following line in a FDCC-set specifies the name of a charmap which may be used  
685

686 with the FDCC-set. It has the following syntax, starting in column 1:  
687  
688 "charmap %s\n",<charmap>  
689

690 This keyword gives a hint on which charmaps a FDCC-set is meant to be supported by.  
691 There may be more than one charmap specification useful with a FDCC-set. It is an  
692 application's responsibility to decide what charmap specification is to be used with that  
693 application.

694 The name is one of:  
695

- 696 - the name of a <charmap> registered by the process defined in ISO/IEC 15897,
- 697 - any other name which may be recognized in some local context - not being  
698 recommended as an international specification.

## 700 4.2 LC\_IDENTIFICATION

701 The LC\_IDENTIFICATION category defines properties of the FDCC-set, and which  
702 specification methods the FDCC-set is conforming to. All keywords are mandatory unless  
703 otherwise noted, and the operands are strings. The following keywords are defined:  
704

706 <b>title</b>	Title of the FDCC-set.
707 <b>source</b>	Organization name of provider of the source.
708 <b>address</b>	Organization postal address.
709 <b>contact</b>	Name of contact person. This keyword is optional.
710 <b>email</b>	Electronic mail address of the organization, or contact 711 person.
712 <b>tel</b>	Telephone number for the organization, in international 713 format.
714 <b>fax</b>	Fax number for the organization, in international format.
715 <b>language</b>	Natural language to which the FDCC-set applies, as specified 716 in ISO 639.
717 <b>territory</b>	The geographic extent where the FDCC-set applies (where 718 applicable), as two-letter form of ISO 3166.
719 <b>audience</b>	If not for general use, an indication of the intended user 720 audience. This keyword is optional.
721 <b>application</b>	If for use of a special application, a description of the 722 application. This keyword is optional.
723 <b>abbreviation</b>	Short name for provider of the source. This keyword is 724 optional.
725 <b>revision</b>	Revision number consisting of digits and zero or more full 726 stops (".").
727 <b>date</b>	Revision date in the format according to this example: 728 "1995-02-05" meaning the 5th of February, 1995.

729 If information required for any of the mandatory keywords above is not available, then the  
730 corresponding string is an empty string. If required information is not present in ISO 639  
731 or ISO 3166, the relevant Maintenance Authority should be approached to get the needed  
732 item registered.

733 Note: Only one language per territory can be addressed with a single FDCC-set; an  
734 additional FDCC-set is required for each additional language for that territory.

<b>category</b>	Is used to define that a category is present and what specification the category is claiming conformance to. The first operand is a string in double-quotes that describes the specification that the category is claiming conformance to, and the following values are defined: "i18n:2000" "posix:1993" The second operand is a string with the category name, where the category names of clause 4 are defined. More than one "category" keyword may be given, but only one per category name.
-----------------	--

The "i18n" LC IDENTIFICATION category is:

```
LC_IDENTIFICATION
% This is the ISO/IEC TR 14652 "i18n" definition for
% the LC_IDENTIFICATION category.
%
title          "ISO/IEC TR 14652 i18n FDCC-set"
source         "ISO/IEC Copyright Office"
address        "Case postale 56, CH-1211 Geneve 20, Switzerland"
contact        ""
email          ""
tel            ""
fax            ""
language       ""
territory      ""
revision       "1.0"
date           "2000-11-15"
%
category      "i18n:2000";LC_IDENTIFICATION
category      "i18n:2000";LC_CTYPE
category      "i18n:2000";LC_COLLATE
category      "i18n:2000";LC_TIME
category      "i18n:2000";LC_NUMERIC
category      "i18n:2000";LC_MONETARY
category      "i18n:2000";LC_MESSAGES
category      "i18n:2000";LC_NAME
category      "i18n:2000";LC_ADDRESS
category      "i18n:2000";LC_TELEPHONE

END LC IDENTIFICATION
```

### 4.3 LC CTYPE

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

A series of characters in a specification can be represented by the hexadecimal symbolic ellipsis symbol ".." (two dots), the decimal symbolic ellipses symbols "...." (4 dots), the double increment hexadecimal symbolic ellipses "..(2)..", or the absolute ellipses "..." (3 dots).

The **hexadecimal symbolic ellipsis** ("..") specification is only valid between symbolic character names. The symbolic names consists of zero or more nonnumeric characters from the set shown with visible glyphs in Table 1, followed by an integer formed by one or more hexadecimal digits, using uppercase letters only for the range "A" to "F". The characters preceding the hexadecimal integer are identical in the two symbolic names, and

798 the integer formed by the hexadecimal digits in the second symbolic name are identical to  
799 or greater than the integer formed by the hexadecimal digits in the first name. This is  
800 interpreted as a series of symbolic names formed from the common part and each of the  
801 integers in hexadecimal format using uppercase letters only between the first and the  
802 second integer, inclusive, and with a length of the symbolic names generated that is equal  
803 to the length of the first (and also the second) symbolic name. As an example,  
804 <U010E>..<U0111> is interpreted as the symbolic names <U010E>, <U010F>, <U0110>,  
805 and <U0111>, in that order.

806  
807 The **decimal symbolic ellipsis** ("....") specification is only valid between symbolic  
808 character names. The symbolic names consist of zero or more nonnumeric characters from  
809 the set shown with visible glyphs in Table 1, followed by an integer formed by one or  
810 more decimal digits. The characters preceding the decimal integer are identical in the two  
811 symbolic names, and the integer formed by the decimal digits in the second symbolic  
812 name is identical to or greater than the integer formed by the decimal digits in the first  
813 name. This is interpreted as a series of symbolic names formed from the common part and  
814 each of the integers in decimal format between the first and the second integer, inclusive,  
815 and with a length of the symbolic names generated that is equal to the length of the first  
816 (and also the second) symbolic name. As an example, <j0101>....<j0104> is interpreted as  
817 the symbolic names <j0101>, <j0102>, <j0103>, and <j0104>, in that order.

818  
819 The **double increment hexadecimal symbolic ellipses** ("..(2)..") works like the  
820 hexadecimal symbolic ellipses, but generates only every other of the symbolic character  
821 names. As an example. <U01AC>..(2)..<U01B2> is interpreted as the symbolic character  
822 names <U01AC>, <U01AE>, <U01B0>, and <U01B2>, in that order.

823  
824 The **absolute ellipsis** specification is only valid within a single encoded character set. An  
825 ellipsis is interpreted as including in the list all characters with an encoded value higher  
826 than the encoded value of the character preceding the ellipsis and lower than the encoded  
827 value of the character following the ellipsis. The absolute ellipsis specification is  
828 deprecated, as this is only relevant to FDCC-sets not using symbolic characters.  
829 As an example, \x30;...;\x39 includes in the character class all characters with encoded  
830 values between the endpoints.

#### 831 4.3.1 Character classification keywords

832  
833 The following keywords are recognized. In the descriptions, the term "automatically  
834 included" means that it is not an error to either include the referenced characters or to  
835 omit them; the interpreting system provides them if missing and accept them silently if  
836 present.

837 838 839 840 841 842 843 844 845 846 847 848 849	<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 is specified.  <b>upper</b> Define characters to be classified as uppercase letters. No character specified for the keywords "cntrl", "digit", "punct", or "space" is specified. The uppercase letters A through Z of the portable character set, automatically belong to this class, with application-defined character values. The keyword may be omitted.  <b>lower</b> Define characters to be classified as lowercase letters. No character specified for the keywords "cntrl", "digit", "punct", or "space" is specified. The lowercase letters a through z of the portable character set, automatically
---	--

850		belong to this class, with application-defined character values. The keyword may be omitted.
851		
852	<b>alpha</b>	Define characters to be classified as used to spell out the words for natural languages; such as letters, syllabic or ideographic characters. No character specified for the keywords "cntrl", "digit", "punct", or "space" is specified. In addition, characters classified as either "upper" or "lower" automatically belong to this class. The keyword may be omitted.
853		
854		
855		
856		
857	<b>digit</b>	Define the characters to be classified as numeric digits. Digits corresponding to the values 0, 1, 2, 3, 4, 5, 6, 7, 8, and 9 can be specified in groups of 10 digits, and in ascending order of the values they represent. The digits of the portable character set are automatically included. If this keyword is not specified, the digits 0 through 9 of the portable character set automatically belong to this class, with application-defined character values. The "digit" keyword is used to specify which characters are accepted as digits in input to an application, such as characters typed in or scanned in from an input text file, and should list digits used with all the scripts supported by the FDCC-set. The keyword may be omitted.
858		
859		
860		
861		
862		
863		
864		
865		
866		
867	<b>alnum</b>	Define the characters to be classified as used to spell out the words for natural languages, and numeric digits. The characters of the "alpha" and "digits" classes are automatically included in this class. The keyword may be omitted.
868		
869		
870		
871	<b>outdigit</b>	Define the characters to be classified as numeric digits for output from an application, such as to a printer or a display or a output text file. Digits corresponding to the values <0>, <1>, <2>, <3>, <4>, <5>, <6>, <7>, <8>, and <9> can be specified, and in ascending order of the values they represent. The intended use is for all places where digits are used for output, including numeric and monetary formatting, and date and time formatting. Only one set of 10 digits may be specified. If this keyword is not specified, the digits 0 through 9 of the portable character set automatically belong to this class, with application-defined character values. The keyword may be omitted.
872		
873		
874		
875		
876		
877		
878		
879		
880		
881	<b>blank</b>	Define characters to be classified as "blank" characters. If this keyword is unspecified, the characters <space> and <tab>, with application-defined character values, belong to this character class.
882		
883		
884	<b>space</b>	Define characters to be classified as white-space characters, to find syntactical boundaries. No character specified for the keywords "upper", "lower", "alpha", "digit", "graph", or "xdigit" is specified. If this keyword is not specified, the characters <space>, <form-feed>, <newline>, <carriage-return>, <tab>, and <vertical-tab>, automatically belong to this class, with application-defined character values. Any characters included in the class "blank" are automatically included. The class should not include the NO-BREAK spaces characters <U00A0>, <U2007>, <UFEFF>, as these characters should not be used for word boundaries. The keyword may be omitted.
885		
886		
887		
888		
889		
890		
891		
892		
893		
894	<b>cntrl</b>	Define characters to be classified as control characters. No character specified for the keywords "upper", "lower", "alpha", "digit", "punct", "graph", "print", or "xdigit" is specified. The keyword is specified.
895		
896		
897	<b>punct</b>	Define characters to be classified as punctuation characters. No character specified for the keywords "upper", "lower", "alpha", "digit", "cntrl", "xdigit", or as the <space> character is specified. The keyword is specified.
898		
899		
900	<b>xdigit</b>	Define the characters to be classified as hexadecimal digits. Only the characters defined for the class "digit" are specified, in ascending sequence
901		

902		by numerical value, followed by sets of six characters representing the hexadecimal digits 10 through 15 in ascending order (for example <A>, <B>, <C>, <D>, <E>, <F>, <a>, <b>, <c>, <d>, <e>, <f>). If this keyword is not specified, the digits <0> through <9>, the uppercase letters "A" through <F>, and the lowercase letters <a> through <f>, automatically belong to this class, with application-defined character values.
903		
904		
905		
906		
907		
908	<b>graph</b>	Define characters to be classified as printable characters, not including the <space> character. If this keyword is not specified, characters specified for the keywords "upper", "lower", "alpha", "digit", "xdigit", and "punct" belong to this character class. No character specified for the keyword "cntrl" is specified.
909		
910		
911		
912		
913	<b>print</b>	Define characters to be classified as printable characters, including the <space> character. If this keyword is not provided, characters specified for the keywords upper, lower, alpha, digit, xdigit, punct, graph, and the <space> character belong to this character class. No character specified for the keyword "cntrl" is specified.
914		
915		
916		
917		
918	<b>toupper</b>	Define the mapping of lowercase letters to uppercase letters. The operand consists of character pairs, separated by semicolons. The characters in each character pair are separated by a comma and the pair enclosed by parentheses. The first character in each pair is the lowercase letter, the second the corresponding uppercase letter. Only characters specified for the keywords "lower" and "upper" are specified. If this keyword is not specified, the lowercase letters <a> through <z>, and their corresponding uppercase letters <A> through <Z>, are automatically included, with application-defined character values.
919		
920		
921		
922		
923		
924		
925		
926		
927	<b>tolower</b>	Define the mapping of uppercase letters to lowercase letters. The operand consists of character pairs, separated by semicolons. The characters in each character pair are separated by a comma and the pair enclosed by parentheses. The first character in each pair is the uppercase letter, the second the corresponding lowercase letter. Only characters specified for the keywords "lower" and "upper" are specified. If this keyword is specified, the uppercase letters <A> through <Z>, and their corresponding lowercase letter, are specified. If this keyword is not specified, the mapping is the reverse mapping of the one specified for toupper.
928		
929		
930		
931		
932		
933		
934		
935		
936	<b>class</b>	Define characters to be classified in the class with the name given in the first operand, which is a string. This string only contains characters of the portable character set that either has the string "LETTER" in its description, or is a digit or <hyphen-minus> or <low-line>. The following operands are characters. This keyword is optional. The keyword can only be specified once per named class. The following two names are recognized:
937		
938		
939		
940		
941		
942		
943		
944		
945		
946		
947		
948		
949		
950	<b>width</b>	Characters to form composite graphic symbols, such as characters listed in ISO/IEC 10646:1993 annex B.1.
951		
952		
953		

width-lists separated by <semicolon> may be given for the various widths. The default value of width of characters in class "cntrl" and class "combining" is 0, else the default value of width is 1. A width for a character may be overridden by a WIDTH specification in a charmap. This keyword is optional.

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

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

Example of use of the "map" keyword:

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

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

Table 2 shows the allowed character class combinations.

## Table 2: Valid Character Class Combinations

Class	upper	lower	alpha	digit	space	cntrl	punct	graph	print	xdigit	blank
upper	+	A	X	X	X	X	A	A	A	+	X
lower	+		A	X	X	X	A	A	A	+	X
alpha	+	+		X	X	X	A	A	A	+	X
digit	X	X	X		X	X	A	A	A	A	X
space	X	X	X		+	*	*	*	*	X	+
cntrl	X	X	X	X	+		X	X	X	X	+
punct	X	X	X	+	X		A	A	A	X	+
graph	+	+	+	+	X	+		A		+	+
print	+	+	+	+	X	+	+		+	+	+
xdigit	+	+	+	+	X	X	A	A			X
blank	X	X	X	A	+	*	*	*	*	X	

NOTES:

Note 1: Explanation of codes:

A Automatically included; see text

+ Permitted

X Mutually exclusive

\* See note 2

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

### 4.3.2 "i18n" LC\_CTYPE category

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

```

1007
1008
1009
1010
1011     LC_CTYPE
1012     % The following is the ISO/IEC TR 14652 i18n fdcc-set LC_CTYPE category.
1013     % It covers ISO/IEC 10646-1 including Cor.1 and AMD 1 thru 9
1014     %
1015     % The "upper" class reflects the uppercase characters of class "alpha"
1016     upper /
1017     % TABLE 1 BASIC LATIN/
1018     <U0041>..<U005A>;/
1019     % TABLE 2 LATIN-1 SUPPLEMENT/
1020     <U00C0>..<U00D6>;<U00D8>..<U00DE>;/
1021     % TABLE 3 LATIN EXTENDED-A/
1022     <U0100>..(2)..<U0136>;/
1023     <U0139>..(2)..<U0147>;/
1024     <U014A>..(2)..<U0178>;/
1025     <U0179>..(2)..<U017D>;/
1026     % TABLE 4 LATIN EXTENDED-B/
1027     <U0181>;<U0182>..(2)..<U0186>;<U0187>;/
1028     <U0189>..<U018B>;<U018E>..<U0191>;<U0193>;<U0194>;/
1029     <U0196>..<U0198>;<U019C>;<U019D>;<U019F>;/
1030     <U01A0>..(2)..<U01A4>;/
1031     <U01A7>;<U01A9>;<U01AC>;<U01AE>;<U01AF>;<U01B1>..<U01B3>;/
1032     <U01B5>;<U01B7>;<U01B8>;<U01BC>;<U01C4>;<U01C5>;<U01C7>;<U01C8>;/
1033     <U01CA>;<U01CB>;/
1034     <U01CD>..(2)..<U01DB>;/
1035     <U01DE>..(2)..<U01EE>;/
1036     <U01F1>;<U01F2>;<U01F4>;<U01FA>..(2)..<U01FE>;/
1037     % TABLE 5 LATIN EXTENDED-B/
1038     <U0200>..(2)..<U0216>;/
1039     % TABLE 6 IPA EXTENSIONS/
1040     % TABLE 9 BASIC GREEK/
1041     <U0386>;<U0388>..<U038A>;<U038C>;<U038E>;<U038F>;<U0391>..<U03A1>;/
1042     <U03A3>..<U03AB>;<U03D2>..<U03D4>;
1043     % TABLE 10 GREEK SYMBOLS AND COPTIC/
1044     <U03E2>..(2)..<U03EE>;/
1045     % TABLE 11 CYRILLIC/
1046     <U0401>..<U040C>;<U040E>..<U042F>;<U0460>..(2)..<U047E>;/
1047     % TABLE 12 CYRILLIC/
1048     <U0480>;<U0490>..(2)..<U04BE>;<U04C1>;<U04C3>;<U04C7>;<U04CB>;/
1049     <U04D0>..(2)..<U04EA>;<U04EE>..(2)..<U04F4>;<U04F8>;/
1050     % TABLE 13 ARMENIAN/
1051     <U0531>..<U0556>;/
1052     % TABLE 28 GEORGIAN/
1053     <U10A0>..<U10C5>;/
1054     % TABLE 31 LATIN EXTENDED ADDITIONAL/
1055     <U1E00>..(2)..<U1E7E>;/
1056     % TABLE 32 LATIN EXTENDED ADDITIONAL/
1057     <U1E80>..(2)..<U1E94>;/
1058     <U1EA0>..(2)..<U1EF8>;/
1059     % TABLE 33 GREEK EXTENDED/
1060     <U1F08>..<U1F0F>;<U1F18>..<U1F1D>;<U1F28>..<U1F2F>;<U1F38>..<U1F3F>;/
1061     <U1F48>..<U1F4D>;<U1F59>..(2)..<U1F5F>;<U1F68>..<U1F6F>;/
1062     % TABLE 34 GREEK EXTENDED/
1063     <U1F88>..<U1F8F>;<U1F98>..<U1F9F>;<U1FA8>..<U1FAF>;<U1FB8>..<U1FBC>;/
1064     <U1FC8>..<U1FCC>;<U1FD8>..<U1FDB>;<U1FE8>..<U1FEC>;<U1FF8>..<U1FFC>
1065     % TABLE 28 GEORGIAN is not addressed as the letters does not have
1066           a uppercase/lowercase relation
1067     %
1068     % The "lower" class reflects the lowercase characters of class "alpha"
1069     lower /
1070     % TABLE 1 BASIC LATIN/
1071     <U0061>..<U007A>;/
1072     % TABLE 2 LATIN-1 SUPPLEMENT/
1073     <U00DF>..<U00F6>;<U00F8>..<U00FF>;/
1074     % TABLE 3 LATIN EXTENDED-A/
1075     <U0101>..(2)..<U0137>;<U0138>..(2)..<U0148>;/
1076     <U0149>..(2)..<U0177>;<U017A>..(2)..<U017E>;<U017F>;/
1077     % TABLE 4 LATIN EXTENDED-B/
1078     <U0180>;<U0183>;<U0185>;<U0188>;<U018C>;<U018D>;<U0192>;<U0195>;/
1079     <U0199>..<U019B>;<U019E>;<U01A1>;<U01A3>;<U01A5>;<U01A8>;<U01AB>;<U01AD>;/
1080     <U01B0>;<U01B4>;<U01B6>;<U01B9>;<U01BA>;<U01BD>;<U01C5>;<U01C6>;/
1081     <U01C8>;<U01C9>;<U01CB>;<U01CC>..(2)..<U01DC>;/
1082     <U01DD>..(2)..<U01F1>;<U01F3>;<U01F5>;<U01FB>;<U01FD>;<U01FF>;/

```

```

1083 % TABLE 5 LATIN EXTENDED-B/
1084 <U0201>..(2)..<U0217>;/
1085 % TABLE 6 IPA EXTENSIONS/
1086 <U0250>..<U0293>;<U0299>..<U02A0>;<U02A3>..<U02A8>;/
1087 % TABLE 9 BASIC GREEK/
1088 <U0390>;<U03AC>..<U03CE>;/
1089 % TABLE 10 GREEK SYMBOLS AND COPTIC/
1090 <U03E2>..(2)..<U03EE>;/
1091 % TABLE 11 CYRILLIC/
1092 <U0430>..<U044F>;<U0451>..<U045C>;<U045E>;<U045F>;<U0460>..(2)..<U047F>;/
1093 % TABLE 12 CYRILLIC/
1094 <U04801>;<U0490>..(2)..<U04BF>;<U04C2>;<U04C4>;<U04C8>;<U04CC>;/
1095 <U04D1>..(2)..<U04EB>;<U04EF>..(2)..<U04F5>;<U04F9>;/
1096 % TABLE 13 ARMENIAN/
1097 <U0561>..<U0587>;/
1098 % TABLE 28 GEORGIAN/
1099 <U10D0>..<U10F6>;/
1100 % TABLE 31 and 32 LATIN EXTENDED ADDITIONAL/
1101 <U1E01>..(2)..<U1E95>;<U1EA1>..(2)..<U1EF9>;/
1102 % TABLE 33 and 34 GREEK EXTENDED/
1103 <U1F08>..<U1F0F>;<U1F18>..<U1F1D>;<U1F28>..<U1F2F>;<U1F38>..<U1F3F>;/
1104 <U1F48>..<U1F4D>;<U1F59>..(2)..<U1F5F>;<U1F68>..<U1F6F>;/
1105 % TABLE 34 GREEK EXTENDED/
1106 <U1F00>..<U1F07>;<U1F10>..<U1F15>;<U1F20>..<U1F27>;<U1F30>..<U1F37>;/
1107 <U1F40>..<U1F45>;<U1F50>..<U1F57>;<U1F60>..<U1F67>;<U1F70>..<U1F7D>;/
1108 <U1F80>..<U1F87>;<U1F90>..<U1F97>;<U1FA0>..<U1FA7>;<U1FB0>..<U1FB4>;/
1109 <U1FB6>;<U1FB7>;<U1FC2>..<U1FC4>;<U1FC6>;<U1FC7>;<U1FD0>..<U1FD3>;/
1110 <U1FD6>;<U1FD7>;<U1FE0>..<U1FE7>;<U1FF2>..<U1FF4>;<U1FF6>;<U1FF7>;/
1111 % TABLE 35 SUPERSCRIPTS AND SUBSCRIPTS, CURRENCY SYMBOLS/
1112 <U207F>
1113 %
1114 % The "alpha" class of the "i18n" FDCC-set is reflecting
1115 % the recommendations in TR 10176 annex A
1116 alpha /
1117 % TABLE 1 BASIC LATIN/
1118 <U0041>..<U005A>;<U0061>..<U007A>;/
1119 % TABLE 2 LATIN-1 SUPPLEMENT/
1120 <U00AA>;<U00BA>;<U00C0>..<U00D6>;<U00D8>..<U00F6>;<U00F8>..<U00FF>;/
1121 % TABLE 3 LATIN EXTENDED-A/
1122 <U0100>..<U017F>;/
1123 % TABLE 4 and 5 LATIN EXTENDED-B/
1124 <U0180>..<U01F5>;<U01FA>..<U0217>;/
1125 % TABLE 6 IPA EXTENSIONS/
1126 <U0250>..<U02A8>;/
1127 % TABLE 31 and 32 LATIN EXTENDED ADDITIONAL/
1128 <U1E00>..<U1E9B>;<U1EA0>..<U1EF9>;/
1129 % TABLE 35 SUPERSCRIPTS AND SUBSCRIPTS, CURRENCY SYMBOLS/
1130 <U207F>;/
1131 % TABLE 9 BASIC GREEK/
1132 <U0386>;<U0388>..<U038A>;<U038C>;<U038E>..<U03A1>;<U03A3>..<U03CE>;/
1133 % TABLE 10 GREEK SYMBOLS AND COPTIC/
1134 <U03D0>..<U03D6>;<U03DA>;<U03DC>;<U03DE>;<U03E0>;<U03E2>..<U03F3>;/
1135 % TABLE 33 and 34 GREEK EXTENDED/
1136 <U1F00>..<U1F15>;<U1F18>..<U1F1D>;<U1F20>..<U1F45>;<U1F48>..<U1F4D>;/
1137 <U1F50>..<U1F57>;<U1F59>;<U1F5B>;<U1F5D>;<U1F5F>..<U1F7D>;/
1138 <U1F80>..<U1FB4>;<U1FB6>..<U1FBC>;<U1FC2>..<U1FC4>;<U1FC6>..<U1FCC>;/
1139 <U1FD0>..<U1FD3>;<U1FD6>..<U1FDB>;<U1FE0>..<U1FEC>;<U1FF2>..<U1FF4>;/
1140 <U1FF6>..<U1FFC>;/
1141 % TABLE 11 and 12 CYRILLIC/
1142 <U0401>..<U040C>;<U040E>..<U044F>;<U0451>..<U045C>;<U045E>..<U0481>;/
1143 <U0490>..<U04C4>;<U04C7>..<U04C8>;<U04CB>..<U04CC>;<U04D0>..<U04EB>;/
1144 <U04EE>..<U04F5>;<U04F8>..<U04F9>;/
1145 % TABLE 13 ARMENIAN/
1146 <U0531>..<U0556>;<U0561>..<U0587>;/
1147 % TABLE 14 HEBREW/
1148 <U05B0>..<U05B9>;<U05BB>..<U05BD>;<U05BF>;<U05C1>..<U05C2>;/
1149 <U05D0>..<U05EA>;<U05F0>..<U05F2>;/
1150 % TABLE 15 and 16 ARABIC/
1151 <U0621>..<U063A>;<U0641>..<U064A>;<U0670>..<U06B7>;<U06BA>..<U06BE>;/
1152 <U06C0>..<U06CE>;<U06D0>..<U06D3>;<U06D5>..<U06DC>;<U06E5>..<U06E8>;/
1153 % TABLE 17 DEVANAGARI/
1154 <U0901>..<U0903>;<U0905>..<U0939>;<U093E>..<U094D>;<U0950>..<U0952>;/
1155 <U0958>..<U0963>;/
1156 % TABLE 18 BENGALI/
1157 <U0981>..<U0983>;<U0985>..<U098C>;<U098F>..<U0990>;/
1158 <U0993>..<U09A8>;<U09AA>..<U09B0>;<U09B2>;<U09B6>..<U09B9>;/
1159 <U09BE>..<U09C4>;<U09C7>..<U09C8>;<U09CB>..<U09CD>;<U09DC>..<U09DD>;/
1160 <U09DF>..<U09E3>;<U09F0>..<U09F1>;/

```

```

1161 % TABLE 19 GURMUKHI/
1162 <U0A02>;<U0A05>..<U0A0A>;<U0A0F>..<U0A10>;<U0A13>..<U0A28>;/
1163 <U0A2A>..<U0A30>;<U0A32>..<U0A33>;<U0A35>..<U0A36>;<U0A38>..<U0A39>;/
1164 <U0A3E>..<U0A42>;<U0A47>..<U0A48>;<U0A4B>..<U0A4D>;<U0A59>..<U0A5C>;/
1165 <U0A5E>;<U0A74>;/
1166 % TABLE 20 GUJARATI/
1167 <U0A81>..<U0A83>;<U0A85>..<U0A8B>;<U0A8D>;<U0A8F>..<U0A91>;/
1168 <U0A93>..<U0AA8>;<U0AAA>..<U0AB0>;<U0AB2>..<U0AB3>;<U0AB5>..<U0AB9>;/
1169 <U0ABD>..<U0AC5>;<U0AC7>..<U0AC9>;<U0ACB>..<U0ACD>;<U0AD0>;<U0AE0>;/
1170 % TABLE 21 ORIYA/
1171 <U0B01>..<U0B03>;<U0B05>..<U0B0C>;<U0B0F>..<U0B10>;<U0B13>..<U0B28>;/
1172 <U0B2A>..<U0B30>;<U0B32>..<U0B33>;<U0B36>..<U0B39>;<U0B3E>..<U0B43>;/
1173 <U0B47>..<U0B48>;<U0B4B>..<U0B4D>;<U0B5C>..<U0B5D>;<U0B5F>..<U0B61>;/
1174 % TABLE 22 TAMIL/
1175 <U0B82>..<U0B83>;<U0B85>..<U0B8A>;<U0B8E>..<U0B90>;<U0B92>..<U0B95>;/
1176 <U0B99>..<U0B9A>;<U0B9C>;<U0B9E>..<U0B9F>;<U0BA3>..<U0BA4>;/
1177 <U0BA8>..<U0BAA>;<U0BAE>..<U0BB5>;<U0BB7>..<U0BB9>;<U0BBE>..<U0BC2>;/
1178 <U0BC6>..<U0BC8>;<U0BCA>..<U0BCD>;/
1179 % TABLE 23 TELUGU/
1180 <U0C01>..<U0C03>;<U0C05>..<U0C0C>;<U0C0E>..<U0C10>;<U0C12>..<U0C28>;/
1181 <U0C2A>..<U0C33>;<U0C35>..<U0C39>;<U0C3E>..<U0C44>;<U0C46>..<U0C48>;/
1182 <U0C4A>..<U0C4D>;<U0C60>..<U0C61>;/
1183 % TABLE 24 KANNADA/
1184 <U0C82>..<U0C83>;<U0C85>..<U0C8C>;<U0C8E>..<U0C90>;<U0C92>..<U0CA8>;/
1185 <U0CAA>..<U0CB3>;<U0CB5>..<U0CB9>;<U0CBE>..<U0CC4>;<U0CC6>..<U0CC8>;/
1186 <U0CCA>..<U0CCD>;<U0CDE>;<U0CE0>..<U0CE1>;/
1187 % TABLE 25 MALAYALAM/
1188 <U0D02>..<U0D03>;<U0D05>..<U0D0C>;<U0D0E>..<U0D10>;<U0D12>..<U0D28>;/
1189 <U0D2A>..<U0D39>;<U0D3E>..<U0D43>;<U0D46>..<U0D48>;<U0D4A>..<U0D4D>;/
1190 <U0D60>..<U0D61>;/
1191 % TABLE 26 THAI/
1192 <U0E01>..<U0E3A>;<U0E40>..<U0E4E>;/
1193 % TABLE 27 LAO/
1194 <U0E81>..<U0E82>;<U0E84>;<U0E87>..<U0E88>;<U0E8A>;<U0E8D>;/
1195 <U0E94>..<U0E97>;<U0E99>..<U0E9F>;<U0EA1>..<U0EA3>;<U0EA5>;<U0EA7>;/
1196 <U0EEA>..<U0EAB>;<U0EAD>..<U0EAE>;<U0EB0>..<U0EB9>;<U0EBB>..<U0EBD>;/
1197 <U0EC0>..<U0EC4>;<U0EC6>;<U0EC8>..<U0ECD>;<U0EDC>..<U0EDD>;/
1198 % TIBETAN Amendment 6/
1199 <U0F00>;<U0F18>..<U0F19>;<U0F35>;<U0F37>;<U0F39>;<U0F40>..<U0F47>;/
1200 <U0F49>..<U0F69>;/
1201 <U0F71>..<U0F84>;<U0F86>..<U0F8B>;<U0F90>..<U0F95>;<U0F97>;/
1202 <U0F99>..<U0FAD>;<U0FB1>..<U0FB7>;<U0FB9>;/
1203 % TABLE 28 GEORGIAN/
1204 <U10A0>..<U10C5>;<U10D0>..<U10F6>;/
1205 % TABLE 50 HIRAGANA/
1206 <U3041>..<U3093>;<U309B>..<U309C>;/
1207 % TABLE 51 KATAKANA/
1208 <U30A1>..<U30F6>;<U30FB>..<U30FC>;/
1209 % TABLE 52 BOPOMOFO/
1210 <U3105>..<U312C>;/
1211 % CJK unified ideographs/
1212 <U4E00>..<U9FA5>;/
1213 % HANGUL amendment 5/
1214 <UAC00>..<UD7A3>;/
1215 % Miscellaneous/
1216 <U00B5>;<U02B0>..<U02B8>;<U02BB>;<U02BD>..<U02C1>;/
1217 <U02D0>..<U02D1>;<U02E0>..<U02E4>;<U037A>;<U0559>;<U093D>;<U0B3D>;/
1218 <U1FBF>;<U2160>..<U2182>;<U3021>..<U3029>;
1219 %
1220 % The "digit" class of the "i18n" FDCC-set is reflecting
1221 % the recommendations in TR 10176 annex A
1222 digit /
1223 % TABLE 1 BASIC LATIN/
1224 <U0030>..<U0039>;/
1225 % TABLE 15 and 16 ARABIC/
1226 <U0660>..<U0669>;<U06F0>..<U06F9>;/
1227 % TABLE 17 DEVANAGARI/
1228 <U0966>..<U096F>;/
1229 % TABLE 18 BENGALI/
1230 <U09E6>..<U09EF>;/
1231 % TABLE 19 GURMUKHI/
1232 <U0A66>..<U0A6F>;/
1233 % TABLE 20 GUJARATI/
1234 <U0AE6>..<U0AEF>;/
1235 % TABLE 21 ORIYA/
1236 <U0B66>..<U0B6F>;/
1237 % TABLE 22 TAMIL/
1238 <0>;<U0BE7>..<U0BEF>;/

```

```

1239 % TABLE 23 TELUGU/
1240 <U0C66>..<U0C6F>; /
1241 % TABLE 24 KANNADA/
1242 <U0CE6>..<U0CEF>; /
1243 % TABLE 25 MALAYALAM/
1244 <U0D66>..<U0D6F>; /
1245 % TABLE 26 THAI/
1246 <U0E50>..<U0E59>; /
1247 % TABLE 27 LAO/
1248 <U0ED0>..<U0ED9>; /
1249 % TIBETAN Amendment 6/
1250 <U0F20>..<U0F29>
1251 %
1252 outdigit <U0030>..<U0039>
1253 %
1254 space /
1255 % ISO/IEC 6429/
1256 <U0008>;<U000A>..<U000D>; /
1257 % TABLE 1 BASIC LATIN/
1258 <U0020>; /
1259 % TABLE 35 GENERAL PUNCTUATION/
1260 <U2000>..<U2006>;<U2008>..<U200B>; /
1261 % TABLE 50 CJK SYMBOLS AND PUNCTUATION, HIRAGANA/
1262 <U3000>
1263 %
1264 cntrl <U0000>..<U001F>;<U007F>..<U009F>
1265 %
1266 punct /
1267 <U0021>..<U002F>;<U003A>..<U0040>;<U005B>..<U0060>;<U007B>..<U007E>; /
1268 <U00A0>..<U00A9>;<U00AB>..<U00B4>;<U00B6>..<U00B9>;<U00BB>..<U00BF>; /
1269 <U00D7>;<U00F7>; /
1270 <U037E>;<U0482>;<U055A>..<U055F>;<U0589>;<U05BE>;<U05C0>;<U05C3>; /
1271 <U05F3>;<U05F4>;<U060C>;<U061B>;<U061F>;<U0640>;<U064B>..<U0652>; /
1272 <U066A>..<U066D>;<U06D4>;<U06DD>..<U06E1>;<U06E9>..<U06EC>;<U10FB>; /
1273 <U2010>..<U2029>;<U2030>..<U2046>;<U20A0>..<U20AA>;<U2100>..<U210B>; /
1274 <U210D>..<U2110>;<U2112>..<U211B>;<U211D>..<U2127>;<U212A>..<U212C>; /
1275 <U212E>..<U2138>;<U2200>..<U22F1>;<U2300>;<U2302>..<U237A>;<U2400>..<U2424>; /
1276 <U2440>..<U244A>;<U2580>..<U2595>;<U25A0>..<U25EF>;<U2600>..<U2613>; /
1277 <U261A>..<U266F>;<U2701>..<U2704>;<U2706>..<U2709>;<U270C>..<U2727>; /
1278 <U2729>..<U274B>;<U274D>;<U274F>..<U2752>;<U2756>;<U2758>..<U275E>; /
1279 <U2761>..<U2767>;<U3000>..<U3020>;<U3030>;<U3036>;<U3037>;<U303F>;<U3164>; /
1280 <U3190>..<U319F>;<U3200>..<U321C>;<U3220>..<U3243>;<U3260>..<U327B>; /
1281 <U327F>..<U32B0>;<U32C0>..<U32CB>;<U32D0>..<U32FE>;<U3300>..<U3376>; /
1282 <U337B>..<U33DD>;<U33E0>..<U33FE>;<UFD3E>;<UFD3F>;<UFE49>..<UFE52>; /
1283 <UFE54>..<UFE66>;<UFE68>..<UFE6B>;<UFEFF>;<UFF01>..<UFF0F>;<UFF1A>..<UFF20>; /
1284 <UFF3B>..<UFF40>;<UFF5B>..<UFF5E>;<UFF61>..<UFF65>;<UFF70>;<UFF9E>..<UFFA0>; /
1285 <UFFE0>..<UFFE6>;<UFFE8>..<UFFEE>;<UFFFD>
1286 %
1287 graph /
1288 <U0021>..<U007E>;<U00A0>..<U01F5>;<U01FA>..<U0217>; /
1289 <U0250>..<U02A8>;<U02B0>..<U02DE>;<U02E0>..<U0300>..<U0345>; /
1290 <U0360>;<U0361>;<U0374>;<U0375>;<U037A>;<U037E>;<U0384>..<U038A>;<U038C>; /
1291 <U038E>..<U03A1>;<U03A3>..<U03CE>;<U03D0>..<U03D6>;<U03DA>;<U03DC>;<U03DE>; /
1292 <U03E0>;<U03E2>..<U03F3>;<U0401>..<U040C>;<U040E>..<U044F>; /
1293 <U0451>..<U045C>;<U045E>..<U0486>;<U0490>..<U04C4>;<U04C7>;<U04C8>; /
1294 <U04CB>;<U04CC>;<U04D0>..<U04EB>;<U04EE>..<U04F5>;<U04F8>;<U04F9>; /
1295 <U0531>..<U0556>;<U0559>..<U055F>;<U0561>..<U0587>;<U0589>; /
1296 <U0591>..<U05A1>;<U05A3>..<U05AF>;<U05B0>..<U05B9>; /
1297 <U05B9>..<U05C4>;<U05D0>..<U05EA>;<U05F0>..<U05F4>;<U060C>;<U061B>;<U061F>; /
1298 <U0621>..<U063A>;<U0640>..<U0652>;<U0660>..<U066D>;<U0670>..<U06B7>; /
1299 <U06BA>..<U06BE>;<U06C0>..<U06CE>;<U06D0>..<U06ED>;<U06F0>..<U06F9>; /
1300 <U0901>..<U0903>;<U0905>..<U0939>;<U093C>..<U094D>;<U0950>..<U0954>; /
1301 <U0958>..<U0970>;<U0981>..<U0983>;<U0985>..<U098C>;<U098F>;<U0990>; /
1302 <U0993>..<U09A8>;<U09AA>..<U09B0>;<U09B2>;<U09B6>..<U09B9>;<U09BC>; /
1303 <U09BE>..<U09C4>;<U09C7>;<U09C8>;<U09CB>..<U09CD>;<U09D7>;<U09DC>;<U09DD>; /
1304 <U09DF>..<U09E3>..<U09E6>..<U09FA>;<U0A02>;<U0A05>..<U0A0A>;<U0A0F>;<U0A10>; /
1305 <U0A13>..<U0A28>;<U0A2A>..<U0A30>;<U0A32>;<U0A33>;<U0A35>;<U0A36>; /
1306 <U0A38>;<U0A39>;<U0A3C>;<U0A3E>..<U0A42>;<U0A47>;<U0A48>;<U0A4B>..<U0A4D>; /
1307 <U0A59>..<U0A5C>;<U0A5E>;<U0A66>..<U0A74>;<U0A81>..<U0A83>;<U0A85>..<U0A8B>; /
1308 <U0A8D>;<U0A8F>..<U0A91>;<U0A93>..<U0AA8>;<U0AAA>..<U0AB0>; /
1309 <U0AB2>;<U0AB3>;<U0AB5>..<U0AB9>;<U0ABC>..<U0AC5>;<U0AC7>..<U0AC9>; /
1310 <U0ACB>..<U0ACD>;<U0AD0>;<U0AE0>;<U0AE6>..<U0AEF>;<U0B01>..<U0B03>; /
1311 <U0B05>..<U0B0C>;<U0B0F>;<U0B10>;<U0B13>..<U0B28>;<U0B2A>..<U0B30>; /
1312 <U0B32>;<U0B33>;<U0B36>..<U0B39>;<U0B3C>..<U0B43>;<U0B47>;<U0B48>; /
1313 <U0B4B>..<U0B4D>;<U0B56>;<U0B57>;<U0B5C>;<U0B5D>;<U0B5F>..<U0B61>; /
1314 <U0B66>..<U0B70>;<U0B82>;<U0B83>;<U0B85>..<U0B8A>;<U0B8E>..<U0B90>; /
1315 <U0B92>..<U0B95>;<U0B99>;<U0B9A>;<U0B9C>;<U0B9E>;<U0B9F>;<U0BA3>;<U0BA4>; /
1316 <U0BA8>..<U0BAA>;<U0BAE>..<U0BB5>;<U0BB7>..<U0BB9>;<U0BBE>..<U0BC2>; /

```

```

1317 <U0BC6>..<U0BC8>;<U0BCA>..<U0BCD>;<U0BD7>;<U0BE7>..<U0BF2>;<U0C01>..<U0C03>;/
1318 <U0C05>..<U0C0C>;<U0C0E>..<U0C10>;<U0C12>..<U0C28>;<U0C2A>..<U0C33>;/
1319 <U0C35>..<U0C39>;<U0C3E>..<U0C44>;<U0C46>..<U0C48>;<U0C4A>..<U0C4D>;/
1320 <U0C55>;<U0C56>;<U0C60>;<U0C61>;<U0C66>..<U0C6F>;<U0C82>;<U0C83>;/
1321 <U0C85>..<U0C8C>;<U0C8E>..<U0C90>;<U0C92>..<U0CA8>;<U0CAA>..<U0CB3>;/
1322 <U0CB5>..<U0CB9>;<U0CBE>..<U0CC4>;<U0CC6>..<U0CC8>;<U0CCA>..<U0CCD>;/
1323 <U0CD5>;<U0CD6>;<U0CDE>;<U0CE0>;<U0CE1>;<U0CE6>..<U0CEF>;<U0D02>;<U0D03>;/
1324 <U0D05>..<U0D0C>;<U0D0E>..<U0D10>;<U0D12>..<U0D28>;<U0D2A>..<U0D39>;/
1325 <U0D3E>..<U0D43>;<U0D46>..<U0D48>;<U0D4A>..<U0D4D>;<U0D57>;<U0D60>;<U0D61>;/
1326 <U0D66>..<U0D6F>;<U0E01>..<U0E3A>;<U0E3F>..<U0E5B>;<U0E81>;<U0E82>;<U0E84>;/
1327 <U0E87>;<U0E88>;<U0E8A>;<U0E8D>;<U0E94>..<U0E97>;<U0E99>..<U0E9F>;/
1328 <U0EA1>..<U0EA3>;<U0EA5>;<U0EA7>;<U0EEA>;<U0EAB>;<U0EAD>..<U0EB9>;/
1329 <U0EBB>..<U0EBD>;<U0EC0>..<U0EC4>;<U0EC6>;<U0EC8>..<U0ECD>;<U0ED0>..<U0ED9>;/
1330 <U0EDC>;<U0EDD>;/
1331 <U0F00>..<U0F47>;<U0F49>..<U0F69>;<U0F71>..<U0F7F>;/
1332 <U10A0>..<U10C5>;<U10D0>..<U10F6>;<U10FB>;<U1100>..<U1159>;/
1333 <U115F>..<U11A2>;<U11A8>..<U11F9>;<U1E00>..<U1E9B>;<U1EA0>..<U1EF9>;/
1334 <U1F00>..<U1F15>;<U1F18>..<U1F1D>;<U1F20>..<U1F45>;<U1F48>..<U1F4D>;/
1335 <U1F50>..<U1F57>;<U1F59>;<U1F5B>;<U1F5D>;<U1F5F>..<U1F7D>;<U1F80>..<U1FB4>;/
1336 <U1FB6>..<U1FC4>;<U1FC6>..<U1FD3>;<U1FD6>..<U1FDB>;<U1FDD>..<U1FEF>;/
1337 <U1FF2>..<U1FF4>;<U1FF6>..<U1FFE>;<U2000>..<U202E>;<U2030>..<U2046>;/
1338 <U206A>..<U2070>;<U2074>..<U208E>;<U20A0>..<U20AC>;<U20D0>..<U20E1>;/
1339 <U2100>..<U2138>;<U2153>..<U2182>;<U2190>..<U21EA>;<U2200>..<U22F1>;<U2300>;/
1340 <U2302>..<U237A>;<U2400>..<U2424>;<U2440>..<U244A>;<U2460>..<U24EA>;/
1341 <U2500>..<U2595>;<U25A0>..<U25EF>;<U2600>..<U2613>;<U261A>..<U266F>;/
1342 <U2701>..<U2704>;<U2706>..<U2709>;<U270C>..<U2727>;<U2729>..<U274B>;<U274D>;/
1343 <U274F>..<U2752>;<U2756>..<U2758>;<U2761>..<U2767>;<U2776>..<U2794>;/
1344 <U2798>..<U27AF>;<U27B1>..<U27BE>;<U3000>..<U3037>;<U303F>;<U3041>..<U3094>;/
1345 <U3099>..<U309E>;<U30A1>..<U30FE>;<U3105>..<U312C>;<U3131>..<U318E>;/
1346 <U3190>..<U319F>;<U3200>..<U321C>;<U3220>..<U3243>;<U3260>..<U327B>;/
1347 <U327F>..<U32B0>;<U32C0>..<U32CB>;<U32D0>..<U32FE>;<U3300>..<U3376>;/
1348 <U337B>..<U33DD>;<U33E0>..<U33FE>;<UFB00>..<UFB06>;<UFB13>..<UFB17>;/
1349 <UFB1E>..<UFB36>;<UFB38>..<UFB3C>;<UFB3E>;<UFB40>;<UFB41>;<UFB43>;<UFB44>;/
1350 <UFB46>..<UFB81>;<UFBBD3>..<UFD3F>;<UFD50>..<UFD8F>;<UFD92>..<UFDC7>;/
1351 <UFDFO>..<UFDDB>;<UFE20>..<UFE23>;<UFE30>..<UFE44>;<UFE49>..<UFE52>;/
1352 <UFE54>..<UFE66>;<UFE68>..<UFE6B>;<UFE70>..<UFE72>;<UFE74>;<UFE76>..<UFEFC>;/
1353 <UFEF4>;<UFF01>..<UFF5E>;<UFF61>..<UFFBE>;<UFFC2>..<UFFC7>;/
1354 <UFFCA>..<UFFCF>;<UFFD2>..<UFFD7>;<UFFDA>..<UFFDC>;<UFFE0>..<UFFE6>;/
1355 <UFFE8>..<UFFEE>;<UFFFC>;<UFFFD>

% "print" is by default "graph", and the <space> character
xdigit <U0030>..<U0039>;<U0041>..<U0046>;<U0061>..<U0066>
blank <U0008>;<U0020>;<U2000>..<U2006>;<U2008>..<U200B>;<U3000>
toupper /
(<U0061>,<U0041>);(<U0062>,<U0042>);(<U0063>,<U0043>);(<U0064>,<U0044>);/
(<U0065>,<U0045>);(<U0066>,<U0046>);(<U0067>,<U0047>);(<U0068>,<U0048>);/
(<U0069>,<U0049>);(<U006A>,<U004A>);(<U006B>,<U004B>);(<U006C>,<U004C>);/
(<U006D>,<U004D>);(<U006E>,<U004E>);(<U006F>,<U004F>);(<U0070>,<U0050>);/
(<U0071>,<U0051>);(<U0072>,<U0052>);(<U0073>,<U0053>);(<U0074>,<U0054>);/
(<U0075>,<U0055>);(<U0076>,<U0056>);(<U0077>,<U0057>);(<U0078>,<U0058>);/
(<U0079>,<U0059>);(<U007A>,<U005A>);(<U00E0>,<U00C0>);(<U00E1>,<U00C1>);/
(<U00E2>,<U00C2>);(<U00E3>,<U00C3>);(<U00E4>,<U00C4>);(<U00E5>,<U00C5>);/
(<U00E6>,<U00C6>);(<U00E7>,<U00C7>);(<U00E8>,<U00C8>);(<U00E9>,<U00C9>);/
(<U00EA>,<U00CA>);(<U00EB>,<U00CB>);(<U00EC>,<U00CC>);(<U00ED>,<U00CD>);/
(<U00EE>,<U00CE>);(<U00EF>,<U00CF>);(<U00F0>,<U00D0>);(<U00F1>,<U00D1>);/
(<U00F2>,<U00D2>);(<U00F3>,<U00D3>);(<U00F4>,<U00D4>);(<U00F5>,<U00D5>);/
(<U00F6>,<U00D6>);(<U00F8>,<U00D8>);(<U00F9>,<U00D9>);(<U00FA>,<U00DA>);/
(<U00FB>,<U00DB>);(<U00FC>,<U00DC>);(<U00FD>,<U00DD>);(<U00FE>,<U00DE>);/
(<U00FF>,<U0178>);(<U0101>,<U0100>);(<U0103>,<U0102>);(<U0105>,<U0104>);/
(<U0107>,<U0106>);(<U0109>,<U0108>);(<U010B>,<U010A>);(<U010D>,<U010C>);/
(<U010F>,<U010E>);(<U0111>,<U0110>);(<U0113>,<U0112>);(<U0115>,<U0114>);/
(<U0117>,<U0116>);(<U0119>,<U0118>);(<U011B>,<U011A>);(<U011D>,<U011C>);/
(<U011F>,<U011E>);(<U0121>,<U0120>);(<U0123>,<U0122>);(<U0125>,<U0124>);/
(<U0127>,<U0126>);(<U0129>,<U0128>);(<U012B>,<U012A>);(<U012D>,<U012C>);/
(<U012F>,<U012E>);(<U0131>,<U0049>);/
(<U0133>,<U0132>);(<U0135>,<U0134>);(<U0137>,<U0136>);/
(<U013A>,<U0139>);(<U013C>,<U013B>);(<U013E>,<U013D>);(<U0140>,<U013F>);/
(<U0142>,<U0141>);(<U0144>,<U0143>);(<U0146>,<U0145>);(<U0148>,<U0147>);/
(<U014B>,<U014A>);(<U014D>,<U014C>);(<U014F>,<U014E>);(<U0151>,<U0150>);/
(<U0153>,<U0152>);(<U0155>,<U0154>);(<U0157>,<U0156>);(<U0159>,<U0158>);/
(<U015B>,<U015A>);(<U015D>,<U015C>);(<U015F>,<U015E>);(<U0161>,<U0160>);/
(<U0163>,<U0162>);(<U0165>,<U0164>);(<U0167>,<U0166>);(<U0169>,<U0168>);/
(<U016B>,<U016A>);(<U016D>,<U016C>);(<U016F>,<U016E>);(<U0171>,<U0170>);/
(<U0173>,<U0172>);(<U0175>,<U0174>);(<U0177>,<U0176>);(<U017A>,<U0179>);/
(<U017C>,<U017B>);(<U017E>,<U017D>);(<U017F>,<U0053>);(<U0183>,<U0182>);/

```

1395 (<U0185>,<U0184>);(<U0188>,<U0187>);(<U018C>,<U018B>);(<U0192>,<U0191>);/  
 1396 (<U0199>,<U0198>);(<U01A1>,<U01A0>);(<U01A3>,<U01A2>);(<U01A5>,<U01A4>);/  
 1397 (<U01A8>,<U01A7>);(<U01AD>,<U01AC>);(<U01B0>,<U01AF>);(<U01B4>,<U01B3>);/  
 1398 (<U01B6>,<U01B5>);(<U01B9>,<U01B8>);(<U01BD>,<U01BC>);(<U01C5>,<U01C4>);/  
 1399 (<U01C6>,<U01C4>);(<U01C8>,<U01C7>);/  
 1400 (<U01C9>,<U01C7>);(<U01CB>,<U01CA>);(<U01CC>,<U01CA>);/  
 1401 (<U01CE>,<U01CD>);(<U01D0>,<U01CF>);(<U01D2>,<U01D1>);(<U01D4>,<U01D3>);/  
 1402 (<U01D6>,<U01D5>);(<U01D8>,<U01D7>);(<U01DA>,<U01D9>);(<U01DC>,<U01DB>);/  
 1403 (<U01DD>,<U018E>);(<U01DF>,<U01DE>);(<U01E1>,<U01E0>);(<U01E3>,<U01E2>);/  
 1404 (<U01E5>,<U01E4>);(<U01E7>,<U01E6>);(<U01E9>,<U01E8>);(<U01EB>,<U01EA>);/  
 1405 (<U01ED>,<U01EC>);(<U01EF>,<U01EE>);(<U01F2>,<U01F1>);/  
 1406 (<U01F3>,<U01F1>);(<U01F5>,<U01F4>);(<U01FB>,<U01FA>);(<U01FD>,<U01FC>);/  
 1407 (<U01FF>,<U01FE>);(<U0201>,<U0200>);(<U0203>,<U0202>);(<U0205>,<U0204>);/  
 1408 (<U0207>,<U0206>);(<U0209>,<U0208>);(<U020B>,<U020A>);(<U020D>,<U020C>);/  
 1409 (<U020F>,<U020E>);(<U0211>,<U0210>);(<U0213>,<U0212>);(<U0215>,<U0214>);/  
 1410 (<U0217>,<U0216>);(<U0253>,<U0181>);(<U0254>,<U0186>);(<U0256>,<U0189>);/  
 1411 (<U0257>,<U018A>);(<U0259>,<U018F>);(<U025B>,<U0190>);(<U0260>,<U0193>);/  
 1412 (<U0263>,<U0194>);(<U0268>,<U0197>);(<U0269>,<U0196>);(<U026F>,<U019C>);/  
 1413 (<U0280>,<U01A6>);/  
 1414 (<U0272>,<U019D>);(<U0275>,<U019F>);(<U0283>,<U01A9>);(<U0288>,<U01AE>);/  
 1415 (<U028A>,<U01B1>);(<U028B>,<U01B2>);(<U0292>,<U01B7>);(<U03AC>,<U0386>);/  
 1416 (<U03AD>,<U0388>);(<U03AE>,<U0389>);(<U03AF>,<U038A>);(<U03B1>,<U0391>);/  
 1417 (<U03B2>,<U0392>);(<U03B3>,<U0393>);(<U03B4>,<U0394>);(<U03B5>,<U0395>);/  
 1418 (<U03B6>,<U0396>);(<U03B7>,<U0397>);(<U03B8>,<U0398>);(<U03B9>,<U0399>);/  
 1419 (<U03B3>,<U039A>);(<U03BB>,<U039B>);(<U03BC>,<U039C>);(<U03BD>,<U039D>);/  
 1420 (<U03BE>,<U039E>);(<U03BF>,<U039F>);(<U03C0>,<U03A0>);(<U03C1>,<U03A1>);/  
 1421 (<U03C2>,<U03A3>);(<U03C3>,<U03A3>);(<U03C4>,<U03A4>);(<U03C5>,<U03A5>);/  
 1422 (<U03C6>,<U03A6>);(<U03C7>,<U03A7>);(<U03C8>,<U03A8>);(<U03C9>,<U03A9>);/  
 1423 (<U03CA>,<U03AA>);(<U03CB>,<U03AB>);(<U03CC>,<U038C>);(<U03CD>,<U038E>);/  
 1424 (<U03CE>,<U038F>);/  
 1425 (<U03E3>,<U03E2>);(<U03E5>,<U03E4>);(<U03E7>,<U03E6>);(<U03E9>,<U03E8>);/  
 1426 (<U03EB>,<U03EA>);(<U03ED>,<U03EC>);(<U03EF>,<U03EE>);/  
 1427 (<U0430>,<U0410>);(<U0431>,<U0411>);(<U0432>,<U0412>);/  
 1428 (<U0433>,<U0413>);(<U0434>,<U0414>);(<U0435>,<U0415>);(<U0436>,<U0416>);/  
 1429 (<U0437>,<U0417>);(<U0438>,<U0418>);(<U0439>,<U0419>);(<U043A>,<U041A>);/  
 1430 (<U043B>,<U041B>);(<U043C>,<U041C>);(<U043D>,<U041D>);(<U043E>,<U041E>);/  
 1431 (<U043F>,<U041F>);(<U0440>,<U0420>);(<U0441>,<U0421>);(<U0442>,<U0422>);/  
 1432 (<U0443>,<U0423>);(<U0444>,<U0424>);(<U0445>,<U0425>);(<U0446>,<U0426>);/  
 1433 (<U0447>,<U0427>);(<U0448>,<U0428>);(<U0449>,<U0429>);(<U044A>,<U042A>);/  
 1434 (<U044B>,<U042B>);(<U044C>,<U042C>);(<U044D>,<U042D>);(<U044E>,<U042E>);/  
 1435 (<U044F>,<U042F>);(<U0451>,<U0401>);(<U0452>,<U0402>);(<U0453>,<U0403>);/  
 1436 (<U0454>,<U0404>);(<U0455>,<U0405>);(<U0456>,<U0406>);(<U0457>,<U0407>);/  
 1437 (<U0458>,<U0408>);(<U0459>,<U0409>);(<U045A>,<U040A>);(<U045B>,<U040B>);/  
 1438 (<U045C>,<U040C>);(<U045E>,<U040E>);(<U045F>,<U040F>);(<U0461>,<U0460>);/  
 1439 (<U0463>,<U0462>);(<U0465>,<U0464>);(<U0467>,<U0466>);(<U0469>,<U0468>);/  
 1440 (<U046B>,<U046A>);(<U046D>,<U046C>);(<U046F>,<U046E>);(<U0471>,<U0470>);/  
 1441 (<U0473>,<U0472>);(<U0475>,<U0474>);(<U0477>,<U0476>);(<U0479>,<U0478>);/  
 1442 (<U047B>,<U047A>);(<U047D>,<U047C>);(<U047F>,<U047E>);(<U0481>,<U0480>);/  
 1443 (<U0491>,<U0490>);(<U0493>,<U0492>);(<U0495>,<U0494>);(<U0497>,<U0496>);/  
 1444 (<U0499>,<U0498>);(<U049B>,<U049A>);(<U049D>,<U049C>);(<U049F>,<U049E>);/  
 1445 (<U04A1>,<U04A0>);(<U04A3>,<U04A2>);(<U04A5>,<U04A4>);(<U04A7>,<U04A6>);/  
 1446 (<U04A9>,<U04A8>);(<U04AB>,<U04AA>);(<U04AD>,<U04AC>);(<U04AF>,<U04AE>);/  
 1447 (<U04B1>,<U04B0>);(<U04B3>,<U04B2>);(<U04B5>,<U04B4>);(<U04B7>,<U04B6>);/  
 1448 (<U04B9>,<U04B8>);(<U04BB>,<U04BA>);(<U04BD>,<U04BC>);(<U04BF>,<U04BE>);/  
 1449 (<U04C2>,<U04C1>);(<U04C4>,<U04C3>);(<U04C8>,<U04C7>);(<U04CC>,<U04CB>);/  
 1450 (<U04D1>,<U04D0>);(<U04D3>,<U04D2>);(<U04D5>,<U04D4>);(<U04D7>,<U04D6>);/  
 1451 (<U04D9>,<U04D8>);(<U04DB>,<U04DA>);(<U04DD>,<U04DC>);(<U04DF>,<U04DE>);/  
 1452 (<U04E1>,<U04E0>);(<U04E3>,<U04E2>);(<U04E5>,<U04E4>);(<U04E7>,<U04E6>);/  
 1453 (<U04E9>,<U04E8>);(<U04EB>,<U04EA>);(<U04EF>,<U04EE>);(<U04F1>,<U04F0>);/  
 1454 (<U04F3>,<U04F2>);(<U04F5>,<U04F4>);(<U04F9>,<U04F8>);(<U0561>,<U0531>);/  
 1455 (<U0562>,<U0532>);(<U0563>,<U0533>);(<U0564>,<U0534>);(<U0565>,<U0535>);/  
 1456 (<U0566>,<U0536>);(<U0567>,<U0537>);(<U0568>,<U0538>);(<U0569>,<U0539>);/  
 1457 (<U056A>,<U053A>);(<U056B>,<U053B>);(<U056C>,<U053C>);(<U056D>,<U053D>);/  
 1458 (<U056E>,<U053E>);(<U056F>,<U053F>);(<U0570>,<U0540>);(<U0571>,<U0541>);/  
 1459 (<U0572>,<U0542>);(<U0573>,<U0543>);(<U0574>,<U0544>);(<U0575>,<U0545>);/  
 1460 (<U0576>,<U0546>);(<U0577>,<U0547>);(<U0578>,<U0548>);(<U0579>,<U0549>);/  
 1461 (<U057A>,<U054A>);(<U057B>,<U054B>);(<U057C>,<U054C>);(<U057D>,<U054D>);/  
 1462 (<U057E>,<U054E>);(<U057F>,<U054F>);(<U0580>,<U0550>);(<U0581>,<U0551>);/  
 1463 (<U0582>,<U0552>);(<U0583>,<U0553>);(<U0584>,<U0554>);(<U0585>,<U0555>);/  
 1464 (<U0586>,<U0556>);/  
 1465 (<U1E01>,<U1E00>);(<U1E03>,<U1E02>);(<U1E05>,<U1E04>);/  
 1466 (<U1E07>,<U1E06>);(<U1E09>,<U1E08>);(<U1E0B>,<U1E0A>);(<U1E0D>,<U1E0C>);/  
 1467 (<U1E0F>,<U1E0E>);(<U1E11>,<U1E10>);(<U1E13>,<U1E12>);(<U1E15>,<U1E14>);/  
 1468 (<U1E17>,<U1E16>);(<U1E19>,<U1E18>);(<U1E1B>,<U1E1A>);(<U1E1D>,<U1E1C>);/  
 1469 (<U1E1F>,<U1E1E>);(<U1E21>,<U1E20>);(<U1E23>,<U1E22>);(<U1E25>,<U1E24>);/  
 1470 (<U1E27>,<U1E26>);(<U1E29>,<U1E28>);(<U1E2B>,<U1E2A>);(<U1E2D>,<U1E2C>);/  
 1471 (<U1E2F>,<U1E2E>);(<U1E31>,<U1E30>);(<U1E33>,<U1E32>);(<U1E35>,<U1E34>);/  
 1472 (<U1E37>,<U1E36>);(<U1E39>,<U1E38>);(<U1E3B>,<U1E3A>);(<U1E3D>,<U1E3C>);/

```

1473  (<U1E3F>,<U1E3E>);(<U1E41>,<U1E40>);(<U1E43>,<U1E42>);(<U1E45>,<U1E44>);/
1474  (<U1E47>,<U1E46>);(<U1E49>,<U1E48>);(<U1E4B>,<U1E4A>);(<U1E4D>,<U1E4C>);/
1475  (<U1E4F>,<U1E4E>);(<U1E51>,<U1E50>);(<U1E53>,<U1E52>);(<U1E55>,<U1E54>);/
1476  (<U1E57>,<U1E56>);(<U1E59>,<U1E58>);(<U1E5B>,<U1E5A>);(<U1E5D>,<U1E5C>);/
1477  (<U1E5F>,<U1E5E>);(<U1E61>,<U1E60>);(<U1E63>,<U1E62>);(<U1E65>,<U1E64>);/
1478  (<U1E67>,<U1E66>);(<U1E69>,<U1E68>);(<U1E6B>,<U1E6A>);(<U1E6D>,<U1E6C>);/
1479  (<U1E6F>,<U1E6E>);(<U1E71>,<U1E70>);(<U1E73>,<U1E72>);(<U1E75>,<U1E74>);/
1480  (<U1E77>,<U1E76>);(<U1E79>,<U1E78>);(<U1E7B>,<U1E7A>);(<U1E7D>,<U1E7C>);/
1481  (<U1E7F>,<U1E7E>);(<U1E81>,<U1E80>);(<U1E83>,<U1E82>);(<U1E85>,<U1E84>);/
1482  (<U1E87>,<U1E86>);(<U1E89>,<U1E88>);(<U1E8B>,<U1E8A>);(<U1E8D>,<U1E8C>);/
1483  (<U1E8F>,<U1E8E>);(<U1E91>,<U1E90>);(<U1E93>,<U1E92>);(<U1E95>,<U1E94>);/
1484  (<U1E9B>,<U1E60>);/
1485  (<U1EA1>,<U1EA0>);(<U1EA3>,<U1EA2>);(<U1EA5>,<U1EA4>);(<U1EA7>,<U1EA6>);/
1486  (<U1EA9>,<U1EA8>);(<U1EAB>,<U1EAA>);(<U1EAD>,<U1EAC>);(<U1EAF>,<U1EAE>);/
1487  (<U1EB1>,<U1EB0>);(<U1EB3>,<U1EB2>);(<U1EB5>,<U1EB4>);(<U1EB7>,<U1EB6>);/
1488  (<U1EB9>,<U1EB8>);(<U1EBB>,<U1EBA>);(<U1EBD>,<U1EBC>);(<U1EBF>,<U1EBE>);/
1489  (<U1EC1>,<U1EC0>);(<U1EC3>,<U1EC2>);(<U1EC5>,<U1EC4>);(<U1EC7>,<U1EC6>);/
1490  (<U1EC9>,<U1EC8>);(<U1ECB>,<U1ECA>);(<U1ECD>,<U1ECC>);(<U1ECF>,<U1ECE>);/
1491  (<U1ED1>,<U1ED0>);(<U1ED3>,<U1ED2>);(<U1ED5>,<U1ED4>);(<U1ED7>,<U1ED6>);/
1492  (<U1ED9>,<U1ED8>);(<U1EDB>,<U1EDA>);(<U1EDD>,<U1EDC>);(<U1EDF>,<U1EDE>);/
1493  (<U1EE1>,<U1EE0>);(<U1EE3>,<U1EE2>);(<U1EE5>,<U1EE4>);(<U1EE7>,<U1EE6>);/
1494  (<U1EE9>,<U1EE8>);(<U1EEB>,<U1EEA>);(<U1EED>,<U1EEC>);(<U1EEF>,<U1EEE>);/
1495  (<U1EF1>,<U1EF0>);(<U1EF3>,<U1EF2>);(<U1EF5>,<U1EF4>);(<U1EF7>,<U1EF6>);/
1496  (<U1EF9>,<U1EF8>);(<U1F00>,<U1F08>);(<U1F01>,<U1F09>);(<U1F02>,<U1F0A>);/
1497  (<U1F03>,<U1F0B>);(<U1F04>,<U1F0C>);(<U1F05>,<U1F0D>);(<U1F06>,<U1F0E>);/
1498  (<U1F07>,<U1F0F>);(<U1F10>,<U1F18>);(<U1F11>,<U1F19>);(<U1F12>,<U1F1A>);/
1499  (<U1F13>,<U1F1B>);(<U1F14>,<U1F1C>);(<U1F15>,<U1F1D>);(<U1F20>,<U1F28>);/
1500  (<U1F21>,<U1F29>);(<U1F22>,<U1F2A>);(<U1F23>,<U1F2B>);(<U1F24>,<U1F2C>);/
1501  (<U1F25>,<U1F2D>);(<U1F26>,<U1F2E>);(<U1F27>,<U1F2F>);(<U1F30>,<U1F38>);/
1502  (<U1F31>,<U1F39>);(<U1F32>,<U1F3A>);(<U1F33>,<U1F3B>);(<U1F34>,<U1F3C>);/
1503  (<U1F35>,<U1F3D>);(<U1F36>,<U1F3E>);(<U1F37>,<U1F3F>);(<U1F40>,<U1F48>);/
1504  (<U1F41>,<U1F49>);(<U1F42>,<U1F4A>);(<U1F43>,<U1F4B>);(<U1F44>,<U1F4C>);/
1505  (<U1F45>,<U1F4D>);(<U1F51>,<U1F59>);(<U1F53>,<U1F5B>);(<U1F55>,<U1F5D>);/
1506  (<U1F57>,<U1F5F>);(<U1F60>,<U1F68>);(<U1F61>,<U1F69>);(<U1F62>,<U1F6A>);/
1507  (<U1F63>,<U1F6B>);(<U1F64>,<U1F6C>);(<U1F65>,<U1F6D>);(<U1F66>,<U1F6E>);/
1508  (<U1F67>,<U1F6F>);(<U1F70>,<U1FBA>);(<U1F71>,<U1FBBA>);(<U1F72>,<U1FC8>);/
1509  (<U1F73>,<U1FC9>);(<U1F74>,<U1FCA>);(<U1F75>,<U1FCB>);(<U1F76>,<U1FDA>);/
1510  (<U1F77>,<U1FDB>);(<U1F78>,<U1FF8>);(<U1F79>,<U1FF9>);(<U1F7A>,<U1FEA>);/
1511  (<U1F7B>,<U1FEB>);(<U1F7C>,<U1FFA>);(<U1F7D>,<U1FFF>);(<U1F80>,<U1F88>);/
1512  (<U1F81>,<U1F89>);(<U1F82>,<U1F8A>);(<U1F83>,<U1F8B>);(<U1F84>,<U1F8C>);/
1513  (<U1F85>,<U1F8D>);(<U1F86>,<U1F8E>);(<U1F87>,<U1F8F>);(<U1F90>,<U1F98>);/
1514  (<U1F91>,<U1F99>);(<U1F92>,<U1F9A>);(<U1F93>,<U1F9B>);(<U1F94>,<U1F9C>);/
1515  (<U1F95>,<U1F9D>);(<U1F96>,<U1F9E>);(<U1F97>,<U1F9F>);(<U1FA0>,<U1FA8>);/
1516  (<U1FA1>,<U1FA9>);(<U1FA2>,<U1FAA>);(<U1FA3>,<U1FAB>);(<U1FA4>,<U1FAC>);/
1517  (<U1FA5>,<U1FAD>);(<U1FA6>,<U1FAE>);(<U1FA7>,<U1FAF>);(<U1FB0>,<U1FB8>);/
1518  (<U1FB1>,<U1FB9>);(<U1FB3>,<U1FBC>);(<U1FC3>,<U1FCC>);(<U1FD0>,<U1FD8>);/
1519  (<U1FD1>,<U1FD9>);(<U1FE0>,<U1FE8>);(<U1FE1>,<U1FE9>);(<U1FE5>,<U1FEC>);/
1520  (<U1FF3>,<U1FFC>)
1521  tolower /
1522  (<U0041>,<U0061>);(<U0042>,<U0062>);(<U0043>,<U0063>);(<U0044>,<U0064>);/
1523  (<U0045>,<U0065>);(<U0046>,<U0066>);(<U0047>,<U0067>);(<U0048>,<U0068>);/
1524  (<U0049>,<U0069>);(<U004A>,<U006A>);(<U004B>,<U006B>);(<U004C>,<U006C>);/
1525  (<U004D>,<U006D>);(<U004E>,<U006E>);(<U004F>,<U006F>);(<U0050>,<U0070>);/
1526  (<U0051>,<U0071>);(<U0052>,<U0072>);(<U0053>,<U0073>);(<U0054>,<U0074>);/
1527  (<U0055>,<U0075>);(<U0056>,<U0076>);(<U0057>,<U0077>);(<U0058>,<U0078>);/
1528  (<U0059>,<U0079>);(<U005A>,<U007A>);(<U00C0>,<U00E0>);(<U00C1>,<U00E1>);/
1529  (<U00C2>,<U00E2>);(<U00C3>,<U00E3>);(<U00C4>,<U00E4>);(<U00C5>,<U00E5>);/
1530  (<U00C6>,<U00E6>);(<U00C7>,<U00E7>);(<U00C8>,<U00E8>);(<U00C9>,<U00E9>);/
1531  (<U00CA>,<U00EA>);(<U00CB>,<U00EB>);(<U00CC>,<U00EC>);(<U00CD>,<U00ED>);/
1532  (<U00CE>,<U00EE>);(<U00CF>,<U00EF>);(<U00D0>,<U00F0>);(<U00D1>,<U00F1>);/
1533  (<U00D2>,<U00F2>);(<U00D3>,<U00F3>);(<U00D4>,<U00F4>);(<U00D5>,<U00F5>);/
1534  (<U00D6>,<U00F6>);(<U00D8>,<U00F8>);(<U00D9>,<U00F9>);(<U00DA>,<U00FA>);/
1535  (<U00DB>,<U00FB>);(<U00DC>,<U00FC>);(<U00DD>,<U00FD>);(<U00DE>,<U00FE>);/
1536  (<U0178>,<U00FF>);(<U0100>,<U0101>);(<U0102>,<U0103>);(<U0104>,<U0105>);/
1537  (<U0106>,<U0107>);(<U0108>,<U0109>);(<U010A>,<U010B>);(<U010C>,<U010D>);/
1538  (<U010E>,<U010F>);(<U0110>,<U0111>);(<U0112>,<U0113>);(<U0114>,<U0115>);/
1539  (<U0116>,<U0117>);(<U0118>,<U0119>);(<U011A>,<U011B>);(<U011C>,<U011D>);/
1540  (<U011E>,<U011F>);(<U0120>,<U0121>);(<U0122>,<U0123>);(<U0124>,<U0125>);/
1541  (<U0126>,<U0127>);(<U0128>,<U0129>);(<U012A>,<U012B>);(<U012C>,<U012D>);/
1542  (<U012E>,<U012F>);(<U0130>,<U0069>);/
1543  (<U0132>,<U0133>);(<U0134>,<U0135>);(<U0136>,<U0137>);/
1544  (<U0139>,<U013A>);(<U013B>,<U013C>);(<U013D>,<U013E>);(<U013F>,<U0140>);/
1545  (<U0141>,<U0142>);(<U0143>,<U0144>);(<U0145>,<U0146>);(<U0147>,<U0148>);/
1546  (<U014A>,<U014B>);(<U014C>,<U014D>);(<U014E>,<U014F>);(<U0150>,<U0151>);/
1547  (<U0152>,<U0153>);(<U0154>,<U0155>);(<U0156>,<U0157>);(<U0158>,<U0159>);/
1548  (<U015A>,<U015B>);(<U015C>,<U015D>);(<U015E>,<U015F>);(<U0160>,<U0161>);/
1549  (<U0162>,<U0163>);(<U0164>,<U0165>);(<U0166>,<U0167>);(<U0168>,<U0169>);/
1550  (<U016A>,<U016B>);(<U016C>,<U016D>);(<U016E>,<U016F>);(<U0170>,<U0171>);/

```

1551       (<U0172>,<U0173>);(<U0174>,<U0175>);(<U0176>,<U0177>);(<U0179>,<U017A>);/  
 1552       (<U017B>,<U017C>);(<U017D>,<U017E>);(<U0182>,<U0183>);(<U0184>,<U0185>);/  
 1553       (<U0187>,<U0188>);(<U0256>,<U0189>);(<U018B>,<U018C>);(<U018E>,<U01DD>);/  
 1554       (<U0191>,<U0192>);(<U0198>,<U0199>);(<U019F>,<U0275>);/  
 1555       (<U01A0>,<U01A1>);(<U01A2>,<U01A3>);/  
 1556       (<U01A4>,<U01A5>);(<U01A6>,<U0280>);/  
 1557       (<U01A7>,<U01A8>);(<U01AC>,<U01AD>);(<U01AF>,<U01B0>);/  
 1558       (<U01B3>,<U01B4>);(<U01B5>,<U01B6>);(<U01B8>,<U01B9>);(<U01BC>,<U01BD>);/  
 1559       (<U01C4>,<U01C6>);(<U01C5>,<U01C6>);(<U01C7>,<U01C9>);/  
 1560       (<U01C8>,<U01C9>);(<U01CA>,<U01CC>);(<U01CB>,<U01CC>);/  
 1561       (<U01CD>,<U01CE>);(<U01CF>,<U01D0>);(<U01D1>,<U01D2>);/  
 1562       (<U01D3>,<U01D4>);(<U01D5>,<U01D6>);(<U01D7>,<U01D8>);(<U01D9>,<U01DA>);/  
 1563       (<U01DB>,<U01DC>);(<U01DE>,<U01DF>);(<U01E0>,<U01E1>);(<U01E2>,<U01E3>);/  
 1564       (<U01E4>,<U01E5>);(<U01E6>,<U01E7>);(<U01E8>,<U01E9>);(<U01EA>,<U01EB>);/  
 1565       (<U01EC>,<U01ED>);(<U01EE>,<U01EF>);(<U01F1>,<U01F3>);/  
 1566       (<U01F2>,<U01F3>);(<U01F4>,<U01F5>);(<U01FA>,<U01FB>);(<U01FC>,<U01FD>);/  
 1567       (<U01FE>,<U01FF>);(<U0200>,<U0201>);(<U0202>,<U0203>);(<U0204>,<U0205>);/  
 1568       (<U0206>,<U0207>);(<U0208>,<U0209>);(<U020A>,<U020B>);(<U020C>,<U020D>);/  
 1569       (<U020E>,<U020F>);(<U0210>,<U0211>);(<U0212>,<U0213>);(<U0214>,<U0215>);/  
 1570       (<U0216>,<U0217>);(<U0181>,<U0253>);(<U0186>,<U0254>);(<U018A>,<U0257>);/  
 1571       (<U018E>,<U0258>);(<U018F>,<U0259>);(<U0190>,<U025B>);(<U0193>,<U0260>);/  
 1572       (<U0194>,<U0263>);(<U0197>,<U0268>);(<U0196>,<U0269>);(<U019C>,<U026F>);/  
 1573       (<U019D>,<U0272>);(<U01A9>,<U0283>);(<U01AE>,<U0288>);(<U01B1>,<U028A>);/  
 1574       (<U01B2>,<U028B>);(<U01B7>,<U0292>);(<U0386>,<U03AC>);(<U0388>,<U03AD>);/  
 1575       (<U0389>,<U03AE>);(<U038A>,<U03AF>);(<U0391>,<U03B1>);(<U0392>,<U03B2>);/  
 1576       (<U0393>,<U03B3>);(<U0394>,<U03B4>);(<U0395>,<U03B5>);(<U0396>,<U03B6>);/  
 1577       (<U0397>,<U03B7>);(<U0398>,<U03B8>);(<U0399>,<U03B9>);(<U039A>,<U03BA>);/  
 1578       (<U039B>,<U03BB>);(<U039C>,<U03BC>);(<U039D>,<U03BD>);(<U039E>,<U03BE>);/  
 1579       (<U039F>,<U03BF>);(<U03A0>,<U03C0>);(<U03A1>,<U03C1>);(<U03A3>,<U03C3>);/  
 1580       (<U03A4>,<U03C4>);(<U03A5>,<U03C5>);(<U03A6>,<U03C6>);(<U03A7>,<U03C7>);/  
 1581       (<U03A8>,<U03C8>);(<U03A9>,<U03C9>);(<U03AA>,<U03CA>);(<U03AB>,<U03CB>);/  
 1582       (<U038C>,<U03CC>);(<U038E>,<U03CD>);(<U038F>,<U03CE>);(<U03E2>,<U03E3>);/  
 1583       (<U03E4>,<U03E5>);(<U03E6>,<U03E7>);(<U03E8>,<U03E9>);(<U03EA>,<U03EB>);/  
 1584       (<U03EC>,<U03ED>);(<U03EE>,<U03EF>);(<U0410>,<U0430>);/  
 1585       (<U0411>,<U0431>);(<U0412>,<U0432>);(<U0413>,<U0433>);(<U0414>,<U0434>);/  
 1586       (<U0415>,<U0435>);(<U0416>,<U0436>);(<U0417>,<U0437>);(<U0418>,<U0438>);/  
 1587       (<U0419>,<U0439>);(<U041A>,<U043A>);(<U041B>,<U043B>);(<U041C>,<U043C>);/  
 1588       (<U041D>,<U043D>);(<U041E>,<U043E>);(<U041F>,<U043F>);(<U0420>,<U0440>);/  
 1589       (<U0421>,<U0441>);(<U0422>,<U0442>);(<U0423>,<U0443>);(<U0424>,<U0444>);/  
 1590       (<U0425>,<U0445>);(<U0426>,<U0446>);(<U0427>,<U0447>);(<U0428>,<U0448>);/  
 1591       (<U0429>,<U0449>);(<U042A>,<U044A>);(<U042B>,<U044B>);(<U042C>,<U044C>);/  
 1592       (<U042D>,<U044D>);(<U042E>,<U044E>);(<U042F>,<U044F>);(<U0401>,<U0451>);/  
 1593       (<U0402>,<U0452>);(<U0403>,<U0453>);(<U0404>,<U0454>);(<U0405>,<U0455>);/  
 1594       (<U0406>,<U0456>);(<U0407>,<U0457>);(<U0408>,<U0458>);(<U0409>,<U0459>);/  
 1595       (<U040A>,<U045A>);(<U040B>,<U045B>);(<U040C>,<U045C>);(<U040E>,<U045E>);/  
 1596       (<U040F>,<U045F>);(<U0460>,<U0461>);(<U0462>,<U0463>);(<U0464>,<U0465>);/  
 1597       (<U0466>,<U0467>);(<U0468>,<U0469>);(<U046A>,<U046B>);(<U046C>,<U046D>);/  
 1598       (<U046E>,<U046F>);(<U0470>,<U0471>);(<U0472>,<U0473>);(<U0474>,<U0475>);/  
 1599       (<U0476>,<U0477>);(<U0478>,<U0479>);(<U047A>,<U047B>);(<U047C>,<U047D>);/  
 1600       (<U047E>,<U047F>);(<U0480>,<U0481>);(<U0490>,<U0491>);(<U0492>,<U0493>);/  
 1601       (<U0494>,<U0495>);(<U0496>,<U0497>);(<U0498>,<U0499>);(<U049A>,<U049B>);/  
 1602       (<U049C>,<U049D>);(<U049E>,<U049F>);(<U04A0>,<U04A1>);(<U04A2>,<U04A3>);/  
 1603       (<U04A4>,<U04A5>);(<U04A6>,<U04A7>);(<U04A8>,<U04A9>);(<U04AA>,<U04AB>);/  
 1604       (<U04AC>,<U04AD>);(<U04AE>,<U04AF>);(<U04B0>,<U04B1>);(<U04B2>,<U04B3>);/  
 1605       (<U04B4>,<U04B5>);(<U04B6>,<U04B7>);(<U04B8>,<U04B9>);(<U04BA>,<U04BB>);/  
 1606       (<U04BC>,<U04BD>);(<U04BE>,<U04BF>);(<U04C1>,<U04C2>);(<U04C3>,<U04C4>);/  
 1607       (<U04C7>,<U04C8>);(<U04CB>,<U04CC>);(<U04D0>,<U04D1>);(<U04D2>,<U04D3>);/  
 1608       (<U04D4>,<U04D5>);(<U04D6>,<U04D7>);(<U04D8>,<U04D9>);(<U04DA>,<U04DB>);/  
 1609       (<U04DC>,<U04DD>);(<U04DE>,<U04DF>);(<U04E0>,<U04E1>);(<U04E2>,<U04E3>);/  
 1610       (<U04E4>,<U04E5>);(<U04E6>,<U04E7>);(<U04E8>,<U04E9>);(<U04EA>,<U04EB>);/  
 1611       (<U04EE>,<U04EF>);(<U04F0>,<U04F1>);(<U04F2>,<U04F3>);(<U04F4>,<U04F5>);/  
 1612       (<U04F8>,<U04F9>);(<U0531>,<U0561>);(<U0532>,<U0562>);(<U0533>,<U0563>);/  
 1613       (<U0534>,<U0564>);(<U0535>,<U0565>);(<U0536>,<U0566>);(<U0537>,<U0567>);/  
 1614       (<U0538>,<U0568>);(<U0539>,<U0569>);(<U053A>,<U056A>);(<U053B>,<U056B>);/  
 1615       (<U053C>,<U056C>);(<U053D>,<U056D>);(<U053E>,<U056E>);(<U053F>,<U056F>);/  
 1616       (<U0540>,<U0570>);(<U0541>,<U0571>);(<U0542>,<U0572>);(<U0543>,<U0573>);/  
 1617       (<U0544>,<U0574>);(<U0545>,<U0575>);(<U0546>,<U0576>);(<U0547>,<U0577>);/  
 1618       (<U0548>,<U0578>);(<U0549>,<U0579>);(<U054A>,<U057A>);(<U054B>,<U057B>);/  
 1619       (<U054C>,<U057C>);(<U054D>,<U057D>);(<U054E>,<U057E>);(<U054F>,<U057F>);/  
 1620       (<U0550>,<U0580>);(<U0551>,<U0581>);(<U0552>,<U0582>);(<U0553>,<U0583>);/  
 1621       (<U1E02>,<U1E03>);(<U1E04>,<U1E05>);(<U1E06>,<U1E07>);(<U1E08>,<U1E09>);/  
 1622       (<U1E0A>,<U1E0B>);(<U1E0C>,<U1E0D>);(<U1E0E>,<U1E0F>);(<U1E10>,<U1E11>);/  
 1623       (<U1E12>,<U1E13>);(<U1E14>,<U1E15>);(<U1E16>,<U1E17>);(<U1E18>,<U1E19>);/  
 1624       (<U1E1A>,<U1E1B>);(<U1E1C>,<U1E1D>);(<U1E1E>,<U1E1F>);(<U1E20>,<U1E21>);/  
 1625       (<U1E22>,<U1E23>);(<U1E24>,<U1E25>);(<U1E26>,<U1E27>);(<U1E28>,<U1E29>);/  
 1626       (<U1E2A>,<U1E2B>);(<U1E2C>,<U1E2D>);(<U1E2E>,<U1E2F>);(<U1E30>,<U1E31>);/  
 1627       (<U1E32>,<U1E33>);(<U1E34>,<U1E35>);(<U1E36>,<U1E37>);(<U1E38>,<U1E39>);/  
 1628       (<U1E3A>,<U1E3B>);(<U1E3C>,<U1E3D>);(<U1E3E>,<U1E3F>);(<U1E40>,<U1E41>);/

```

1629 (<U1E42>,<U1E43>);(<U1E44>,<U1E45>);(<U1E46>,<U1E47>);(<U1E48>,<U1E49>);/
1630 (<U1E4A>,<U1E4B>);(<U1E4C>,<U1E4D>);(<U1E4E>,<U1E4F>);(<U1E50>,<U1E51>);/
1631 (<U1E52>,<U1E53>);(<U1E54>,<U1E55>);(<U1E56>,<U1E57>);(<U1E58>,<U1E59>);/
1632 (<U1E5A>,<U1E5B>);(<U1E5C>,<U1E5D>);(<U1E5E>,<U1E5F>);(<U1E60>,<U1E61>);/
1633 (<U1E62>,<U1E63>);(<U1E64>,<U1E65>);(<U1E66>,<U1E67>);(<U1E68>,<U1E69>);/
1634 (<U1E6A>,<U1E6B>);(<U1E6C>,<U1E6D>);(<U1E6E>,<U1E6F>);(<U1E70>,<U1E71>);/
1635 (<U1E72>,<U1E73>);(<U1E74>,<U1E75>);(<U1E76>,<U1E77>);(<U1E78>,<U1E79>);/
1636 (<U1E7A>,<U1E7B>);(<U1E7C>,<U1E7D>);(<U1E7E>,<U1E7F>);(<U1E80>,<U1E81>);/
1637 (<U1E82>,<U1E83>);(<U1E84>,<U1E85>);(<U1E86>,<U1E87>);(<U1E88>,<U1E89>);/
1638 (<U1E8A>,<U1E8B>);(<U1E8C>,<U1E8D>);(<U1E8E>,<U1E8F>);(<U1E90>,<U1E91>);/
1639 (<U1E92>,<U1E93>);(<U1E94>,<U1E95>);(<U1EA0>,<U1EA1>);(<U1EA2>,<U1EA3>);/
1640 (<U1EA4>,<U1EA5>);(<U1EA6>,<U1EA7>);(<U1EA8>,<U1EA9>);(<U1EAA>,<U1EAB>);/
1641 (<U1EAC>,<U1EAD>);(<U1EAE>,<U1EAF>);(<U1EB0>,<U1EB1>);(<U1EB2>,<U1EB3>);/
1642 (<U1EB4>,<U1EB5>);(<U1EB6>,<U1EB7>);(<U1EB8>,<U1EB9>);(<U1EBA>,<U1EBB>);/
1643 (<U1EBC>,<U1EBD>);(<U1EBE>,<U1EBF>);(<U1EC0>,<U1EC1>);(<U1EC2>,<U1EC3>);/
1644 (<U1EC4>,<U1EC5>);(<U1EC6>,<U1EC7>);(<U1EC8>,<U1EC9>);(<U1ECA>,<U1ECB>);/
1645 (<U1ECC>,<U1ECD>);(<U1ECE>,<U1ECF>);(<U1ED0>,<U1ED1>);(<U1ED2>,<U1ED3>);/
1646 (<U1ED4>,<U1ED5>);(<U1ED6>,<U1ED7>);(<U1ED8>,<U1ED9>);(<U1EDA>,<U1EDB>);/
1647 (<U1EDC>,<U1EDD>);(<U1EDE>,<U1EDF>);(<U1EE0>,<U1EE1>);(<U1EE2>,<U1EE3>);/
1648 (<U1EE4>,<U1EE5>);(<U1EE6>,<U1EE7>);(<U1EE8>,<U1EE9>);(<U1EEA>,<U1EEB>);/
1649 (<U1EEC>,<U1EED>);(<U1EEE>,<U1EEF>);(<U1EF0>,<U1EF1>);(<U1EF2>,<U1EF3>);/
1650 (<U1EF4>,<U1EF5>);(<U1EF6>,<U1EF7>);(<U1EF8>,<U1EF9>);(<U1F08>,<U1F00>);/
1651 (<U1F09>,<U1F01>);(<U1F0A>,<U1F02>);(<U1F0B>,<U1F03>);(<U1F0C>,<U1F04>);/
1652 (<U1F0D>,<U1F05>);(<U1F0E>,<U1F06>);(<U1F0F>,<U1F07>);(<U1F18>,<U1F10>);/
1653 (<U1F19>,<U1F11>);(<U1F1A>,<U1F12>);(<U1F1B>,<U1F13>);(<U1F1C>,<U1F14>);/
1654 (<U1F1D>,<U1F15>);(<U1F28>,<U1F20>);(<U1F29>,<U1F21>);(<U1F2A>,<U1F22>);/
1655 (<U1F2B>,<U1F23>);(<U1F2C>,<U1F24>);(<U1F2D>,<U1F25>);(<U1F2E>,<U1F26>);/
1656 (<U1F2F>,<U1F27>);(<U1F38>,<U1F30>);(<U1F39>,<U1F31>);(<U1F3A>,<U1F32>);/
1657 (<U1F3B>,<U1F33>);(<U1F3C>,<U1F34>);(<U1F3D>,<U1F35>);(<U1F3E>,<U1F36>);/
1658 (<U1F3F>,<U1F37>);(<U1F48>,<U1F40>);(<U1F49>,<U1F41>);(<U1F4A>,<U1F42>);/
1659 (<U1F4B>,<U1F43>);(<U1F4C>,<U1F44>);(<U1F4D>,<U1F45>);(<U1F59>,<U1F51>);/
1660 (<U1F5B>,<U1F53>);(<U1F5D>,<U1F55>);(<U1F5F>,<U1F57>);(<U1F68>,<U1F60>);/
1661 (<U1F69>,<U1F61>);(<U1F6A>,<U1F62>);(<U1F6B>,<U1F63>);(<U1F6C>,<U1F64>);/
1662 (<U1F6D>,<U1F65>);(<U1F6E>,<U1F66>);(<U1F6F>,<U1F67>);(<U1FBA>,<U1F70>);/
1663 (<U1FBB>,<U1F71>);(<U1FC8>,<U1F72>);(<U1FC9>,<U1F73>);(<U1FCA>,<U1F74>);/
1664 (<U1FCB>,<U1F75>);(<U1FDA>,<U1F76>);(<U1FDB>,<U1F77>);(<U1FF8>,<U1F78>);/
1665 (<U1FF9>,<U1F79>);(<U1FEA>,<U1F7A>);(<U1FEB>,<U1F7B>);(<U1FFA>,<U1F7C>);/
1666 (<U1FFB>,<U1F7D>);(<U1F88>,<U1F80>);(<U1F89>,<U1F81>);(<U1F8A>,<U1F82>);/
1667 (<U1F8B>,<U1F83>);(<U1F8C>,<U1F84>);(<U1F8D>,<U1F85>);(<U1F8E>,<U1F86>);/
1668 (<U1F8F>,<U1F87>);(<U1F98>,<U1F90>);(<U1F99>,<U1F91>);(<U1F9A>,<U1F92>);/
1669 (<U1F9B>,<U1F93>);(<U1F9C>,<U1F94>);(<U1F9D>,<U1F95>);(<U1F9E>,<U1F96>);/
1670 (<U1F9F>,<U1F97>);(<U1FA8>,<U1FA0>);(<U1FA9>,<U1FA1>);(<U1FAA>,<U1FA2>);/
1671 (<U1FAB>,<U1FA3>);(<U1FAC>,<U1FA4>);(<U1FAD>,<U1FA5>);(<U1FAE>,<U1FA6>);/
1672 (<U1FAF>,<U1FA7>);(<U1FB8>,<U1FB0>);(<U1FB9>,<U1FB1>);(<U1FBC>,<U1FB3>);/
1673 (<U1FCC>,<U1FC3>);(<U1FD8>,<U1FD0>);(<U1FD9>,<U1FD1>);(<U1FE8>,<U1FE0>);/
1674 (<U1FE9>,<U1FE1>);(<U1FEC>,<U1FE5>);(<U1FFC>,<U1FF3>);

% The "combining" class reflects ISO/IEC 10646-1 annex B.1
% That is, all combining characters (level 2+3).
class "combining" /
<U0300>..<U036F>; <U20D0>..<U20FF>; <UFE20>..<UFE2F>; /
<U0483>..<U0486>;<U0591>..<U05A1>;<U05A3>..<U05B9>; /
<U05B8>..<U05BD>;<U05BF>;<U05C1>;<U05C2>;<U05C4>;<U064B>..<U0652>;<U0670>; /
<U06D7>..<U06E4>;<U06E7>;<U06E8>;<U06EA>..<U06ED>;<U0901>..<U0903>;<U093C>; /
<U093E>..<U094D>;<U0951>..<U0954>;<U0962>;<U0963>;<U0981>..<U0983>;<U09BC>; /
<U09BE>..<U09C4>;<U09C7>;<U09C8>;<U09CB>..<U09CD>;<U09D7>;<U09E2>;<U09E3>; /
<U0A02>;<U0A3C>;<U0A3E>..<U0A42>;<U0A47>;<U0A48>;<U0A4B>..<U0A4D>; /
<U0A70>;<U0A71>;<U0A81>..<U0A83>;<U0ABC>;<U0ABE>..<U0AC5>;<U0AC7>..<U0AC9>; /
<U0ACB>..<U0ACD>;<U0B01>..<U0B03>;<U0B3C>;<U0B3E>..<U0B43>;<U0B47>;<U0B48>; /
<U0B4B>..<U0B4D>;<U0B56>;<U0B57>;<U0B82>;<U0B83>;<U0B8E>..<U0BC2>; /
<U0BC6>..<U0BC8>;<U0BCA>..<U0BCD>;<U0BD7>;<U0C01>..<U0C03>;<U0C3E>..<U0C44>; /
<U0C46>..<U0C48>;<U0C4A>..<U0C4D>;<U0C55>;<U0C56>;<U0C82>;<U0C83>; /
<U0CBE>..<U0CC4>;<U0CC6>..<U0CC8>;<U0CCA>..<U0CCD>;<U0CD5>;<U0CD6>; /
<U0D02>;<U0D03>;<U0D3E>..<U0D43>;<U0D46>..<U0D48>;<U0D4A>..<U0D4D>;<U0D57>; /
<U0E31>;<U0E34>..<U0E47>..<U0E4E>;<U0EB1>;<U0EB4>..<U0EB9>; /
<U0EBB>;<U0EBCB>;<U0EC8>..<U0ECD>;<U0F18>;<U0F19>;<U0F35>;<U0F37>;<U0F39>; /
<U0F3E>;<U0F3F>;<U0F71>..<U0F84>;<U0F86>..<U0F89>;<U0F8B>;<U0F90>..<U0F95>; /
<U0F97>;<U0F99>..<U0FAD>;<U0FB1>..<U0FB7>;<U0FB9>;<U302A>..<U302F>; /
<U3099>;<U309A>;<UFB1E>

% The "combining_level3" class reflects ISO/IEC 10646-1 annex B.2
% That is, combining characters of level 3.
class "combining_level3"; /
<U0300>..<U036F>;<U20D0>..<U20FF>;<U1100>..<U11FF>;<UFE20>..<UFE2F>; /
<U0483>..<U0486>;<U0591>..<U05A1>;<U05A3>..<U05AE>;<U05C4>; /
<U05AF>;<U093C>;<U0953>;<U0954>;<U09BC>;<U09D7>;<U0A3C>; /
<U0A70>;<U0A71>;<U0ABC>;<U0B3C>;<U0B56>;<U0B57>;<U0BD7>;<U0C55>;<U0C56>; /
<U0CD5>;<U0CD6>;<U0D57>;<U0F39>;<U302A>..<U302F>;<U3099>;<U309A>

```

```

1707 %
1708 width /
1709   <U200B>;<U200C>;<U200D>;<U200E>;<U200F>; <U202A>; <U202B>;/
1710   <U202C>; <U202D>;<U202E>; <UEFF> : 0;/
1711   <U1100>..<U115F>;<U2E80>..<U3009>;<U300C>..<U3019>;/
1712   <U301C>..<U303E>;<U3040>..<UA4CF>;<UAC00>..<UD7A3>;/
1713   <UF900>..<UFAFF>;<UFE30>..<UFE6F>;<UFF00>..<UFF5F>;/
1714   <UFFE0>..<UFFE6> : 2
1715
1716 END LC_CTYPE
1717
1718

```

#### 4.4 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 assign such values, but that ordering of strings using the resultant collation definition in the FDCC-set behaves as if such assignment is done and used in the collation process. The collation sequence definition is used by regular expressions, pattern matching. When no weights are specified the collation sequence definition also is used for sorting, else the weighting defines the 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 is 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 are 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) Easy reordering of characters. ISO/IEC 14651 has a template for collation specification that with just a few modifications can be culturally correct for a specific culture. Here the "reorder-after" keyword gives a convenient way to modify a FDCC-set template.
- (9) Easy reordering of sections. The template in ISO/IEC 14651 gives an ordering of the sections that may not be culturally acceptable in certain cultures. The keyword "reorder-section-after" gives a convenient way to modify the order of sections in a FDCC-set template.

1762	The following keywords are recognized in a collation sequence definition. Some of them 1763 are described in detail in the following subclauses. The keywords are mandatory unless 1764 otherwise noted.
1765	
1766	<b>copy</b> Specify the name of an existing FDCC-set to be used 1767 as the source for the definition of this category. If 1768 this keyword is specified, only the "reorder-after", 1769 "reorder-end", "reorder-section-after" and "reorder- 1770 section-end" keywords may also be specified. The 1771 FDCC-set is copied in source form.
1772	<b>coll_weight_max</b> Define as a decimal number the number of collation 1773 levels that an interpreting system needs to support 1774 for this FDCC-set, this value is elsewhere referred to 1775 as the COLL_WEIGHT_MAX limit (e.g. in the 1776 "order_start" statement). An interpreting system 1777 caters for up to 7 collating levels.
1778	<b>section-symbol</b> Define a section symbol representing a set of 1779 collation order statements. The section is defined 1780 with the "order_start" keyword until the next 1781 "order_start" or "order_end" keyword. This keyword 1782 is optional.
1783	<b>collating-element</b> Define a collating-element symbol representing a 1784 multicharacter collating element. This keyword is 1785 optional.
1786	<b>collating-symbol</b> Define one or more collating symbols for use in 1787 collation order statements. This keyword is optional.
1788	<b>symbol-equivalence</b> Define a collating-symbol to be equivalent to another 1789 defined collating-symbol.
1790	<b>order_start</b> Define collation rules. This statement is followed by 1791 one or more collation order statements, assigning 1792 character collation values and collation weights to 1793 collating elements.
1794	<b>order_end</b> Specify the end of the collation-order statements.
1795	<b>reorder-after</b> Redefine collating rules. Specify after which 1796 collating element the redefinition of collation order 1797 takes order. This statement is followed by one or 1798 more collation order statements, reassigned character 1799 collation values and collation weights to collating 1800 elements.
1801	<b>reorder-end</b> Specify the end of the "reorder-after" collating order 1802 statements.
1803	<b>reorder-section-after</b> Redefine the order of sections. This statement is 1804 followed by one or more section symbols, 1805 reassigned character collation values and collation 1806 weights to collating elements.
1807	<b>reorder-section-end</b> Specify the end of the "reorder-section" section order 1808 statements.
1809	
1810	<b>4.4.1 Collation statements</b>
1811	
1812	The "order_start" and "reorder-after" keywords are followed by collating statements. The 1813 syntax for the collating statements is

1814        "%s %s;%s;...;%s\n",<collating-identifier>,<weight>,<weight>,...  
1815  
1816        Each <collating-identifier> consists of either a character (in any of the forms defined in  
1817        4.1.1), a <collating-element>, a <collating-symbol>, an ellipsis, or the special symbol  
1818        "UNDEFINED". The weights for each of the collation elements determines the character  
1819        collation sequence - such that each collation statement does not need to be in collation  
1820        order, and weights could be rearranged via for example the "reorder-after" keyword. No  
1821        character has any specific predetermined placement in the collation sequence. The order in  
1822        which collating elements are specified determines the character collation sequence, such  
1823        that each collating element compares less than the elements following it.  
1824  
1825        A <collating-element> is used to specify multicharacter collating elements, and indicates  
1826        that the character sequence specified via the <collating-element> is to be collated as a unit  
1827        and in the relative order specified by its place in the list of collating statements.  
1828  
1829        A <collating-symbol> is used to define a position in the relative order for use in weights.  
1830  
1831        The absolute ellipsis symbol ("...") specifies that a sequence of characters collate according  
1832        to their encoded character values. It is interpreted as indicating that all characters with a  
1833        coded character set value higher than the value of the character in the preceding line, and  
1834        lower than the coded character set value for the character in the following line, in the  
1835        current coded character set, are placed in the character collation order between the  
1836        previous and the following character in ascending order according to their coded character  
1837        set values. An initial ellipsis is interpreted as if the preceding line specified the <NUL>  
1838        character, and a trailing ellipsis as if the following line specified the highest coded  
1839        character set value in the current coded character set. An ellipsis is treated as invalid if the  
1840        preceding or following lines do not specify characters in the current coded character set.  
1841        The use of the ellipsis symbol ties the definition to a specific coded character set and may  
1842        preclude the definition from being portable between applications, and is depreciated.  
1843        Symbolic ellipses may be used as the ellipses symbol, but generating symbolic character  
1844        names, and thus have a better chance of portability between applications.  
1845  
1846        The symbolic ellipses (".." or "....") specifies a sequence of collating statements. It is  
1847        interpreted as indicating that all characters with symbolic names higher than the symbolic  
1848        name of the character in the preceding line, and lower in the sequence of symbolic names  
1849        for the character in the following line, is placed in the character collation order between  
1850        the previous and the following character in ascending order.  
1851  
1852        The symbol "UNDEFINED" is interpreted as including all coded character set values not  
1853        specified explicitly or via the ellipsis or one of the symbolic ellipses symbols. Such  
1854        characters are inserted in the character collation order at the point indicated by the symbol,  
1855        and in ascending order according to their coded character set values. If no "UNDEFINED"  
1856        symbol is specified, and the current coded character set contains characters not specified  
1857        in this clause, the utility issues a warning message and place such characters at the end of  
1858        the character collation order.  
1859  
1860        The optional operands for each collation-element are used to define the primary,  
1861        secondary, or subsequent weights for the collating element. The first operand specifies the  
1862        relative primary weight, the second the relative secondary weight, and so on. Two or more  
1863        collation-elements can be assigned the same weight; they belong to the same equivalence  
1864        class if they have the same primary weight. Collation behaves as if, for each weight level,  
1865        "IGNORE"d elements are removed. Then each successive pair of elements is compared

1866 according to the relative weights for the elements. If the two strings compare equal, the  
 1867 process is repeated for the next weight level, up to the limit "COLL\_WEIGHTS\_MAX" of  
 1868 the associated FDCC-set.

1869  
 1870 Weights are expressed as characters (in any of the forms specified here), <collating-  
 1871 symbol>s, <collating-element>s, an ellipsis, or the special symbol "IGNORE". A single  
 1872 character, a <collating-symbol>, or a <collating-element> represent the relative order in  
 1873 the character collating sequence of the character or symbol, rather than the character or  
 1874 characters themselves.

1875  
 1876 One-to-many mapping is indicated by specifying two or more concatenated characters or  
 1877 symbolic names. Thus, if the character <ss> is given the string <s><s> as a weight,  
 1878 comparisons are performed as if all occurrences of the character <ss> are replaced by  
 1879 <s><s>. If it is desirable to define <ss> and <s><s> as an equivalence class, then a  
 1880 collating-element must be defined for the string "ss", as in the example below.

1881  
 1882 All characters specified via an ellipsis are by default assigned unique weights, equal to the  
 1883 relative order of characters. Characters specified via an explicit or implicit "UNDEFINED"  
 1884 special symbol are by default assigned the same primary weight (i.e., belong to the same  
 1885 equivalence class). An ellipsis symbol as a weight is interpreted to mean that each  
 1886 character in the sequence has unique weights, equal to the relative order of their character  
 1887 in the character collation sequence. Secondary and subsequent weights have unique values.  
 1888 The use of the ellipsis as a weight is treated as an error if the collating element is neither  
 1889 an ellipsis nor the special symbol "UNDEFINED".

1890  
 1891 The special keyword "IGNORE" as a weight indicates that when strings are compared  
 1892 using the weights at the level where "IGNORE" is specified, the collating element is  
 1893 ignored; i.e., as if the string did not contain the collating element. In regular expressions  
 1894 and pattern matching, all characters that are "IGNORE"d in their primary weight form an  
 1895 equivalence class.

1896  
 1897 A <comment\_character> occurring where the delimiter ";" may occur, terminates the  
 1898 collating statement.

1899  
 1900 An empty operand is interpreted as the collating-element itself.

1901  
 1902 For example, the collation statement

1903  
 1904   <a>   <a>;<a>

1905  
 1906 is equal to

1907  
 1908   <a>

1909  
 1910 An ellipsis (absolute or symbolic) can be used as an operand if the collating-element was  
 1911 an ellipsis, and is interpreted as the value of each character defined by the ellipsis.

1912  
 1913 Example:

1914  
 1915   collating-element <ch> from "<c><h>"  
 1916   collating-element <Ch> from "<C><h>"  
 1917   order\_start   forward;backward  
 1918   UNDEFINED     IGNORE;IGNORE  
 1919   <LOW>  
 1920   <space>       <LOW>;<space>  
 1921   ...            <LOW>;

```

1922      <a>          <a>;<a>
1923      <a'>         <a>;<a'>
1924      <A>          <a>;<A>
1925      <A'>         <a>;<A'>
1926      <ch>          <ch>;<ch>
1927      <Ch>          <ch>;<Ch>
1928      <s>           <s>;<s>
1929      <ss>          "<s><s>" ; "<ss><ss>""
1930      order_end
1931

```

This example is interpreted as follows:

- 1932 (1) The UNDEFINED means that all characters not specified in this definition (explicitly or via the ellipsis) is ignored.
- 1933 (2) <LOW> defines the first collating weight, and thus the lowest weight in this example.
- 1934 (3) All characters between <space> and <a> have the same primary equivalence class <LOW> and individual secondary weights based on their ordinal encoded values. (The use of absolute ellipses is depreciated, but used here to illustrate generic use of ellipses. Symbolic ellipses should be used instead).
- 1935 (4) All characters based on the upper or lowercase character "a" belong to the same primary equivalence class.
- 1936 (5) The multicharacter collating element <c><h> is represented by the collating symbol <ch> and belongs to the same primary equivalence class as the multicharacter collating element <C><h>.
- 1937 (6) The <ss> collating element has two weights on the primary level, and it is in the same primary equivalence class as two consecutive <s>-es; on the secondary level the collating element has two weights of the equivalence class <ss>.

#### 4.4.2 "copy" keyword

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

```
"copy %s\n", <FDCC-set-name>
```

The <FDCC-set-name> consists of one or more characters (in any of the forms defined in 4.1.1). If this keyword is specified, only the "reorder-after", "reorder-end", "reorder-section-after" and "reorder-section-end" keywords may also be specified. The FDCC-set is copied in source form.

#### 4.4.3 "coll\_weight\_max" keyword

This keyword defines as a decimal number the number of collation levels that an interpreting system needs to support. An interpreting system caters for up to 7 collating levels. The syntax is

```
"coll_weight_max %d\n", <value>
```

#### 4.4.4 "section-symbol" keyword

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

```
"section-symbol %s\n", <section-symbol>
```

The <section-symbol> is a symbolic name, enclosed between angle brackets (< and >), and does not duplicate any symbolic name in the current charmap (if any), or any other symbolic name defined in this collation definition. A <section-symbol> defined via this

1980 keyword is only defined within the LC\_COLLATE category.

1981

1982 Example:  
 1983 section-symbol <LATIN>  
 1984 section-symbol <ARABIC>

1985

#### 1986 4.4.5 "collating-element" keyword

1987

1988 In addition to the collating elements in the character set, the collating-element keyword is  
 1989 used to define multicharacter collating elements. The syntax is

1990

1991 "collating-element %s from %s\n",<collating-symbol>,<string>

1992

1993 The <collating-symbol> operand is a symbolic name, enclosed between angle brackets (<  
 1994 and >), and does not duplicate any symbolic name in the current charmap or repertoiremap  
 1995 file (if any), or any other symbolic name defined in this collation definition. The string  
 1996 operand is a string of two or more characters that collates as an entity. A <collating-  
 1997 element> defined via this keyword is only defined within the LC\_COLLATE category.

1998

1999 Example with ISO/IEC 10646-1:  
 2000 collating-element <ch> from "<c><h>"  
 2001 collating-element <e-acute> from "<e><combining-acute>"  
 2002 collating-element <aa> from "<a><a>"

2003

2004 Note: The problem of comparing a fully composed character of ISO/IEC 10646 with a  
 2005 decomposed representation of the same text is sometimes handled by the two strings  
 2006 comparing equal up to level 3 (the case level) of ISO/IEC 14651, but distinguishing the  
 2007 two at the 4th level.

2008

#### 2009 4.4.6 "collating-symbol" keyword

2010

2011 This keyword is used to define symbols for use in collation sequence statements; e.g.,  
 2012 between the order\_start and the order\_end keywords. The syntax is

2013

2014 "collating-symbol %s;%s;...%s\n", <collating-symbol>, <collating-symbol> ...

2015

2016 The <collating-symbol> is a symbolic name, enclosed between angle brackets (< and >),  
 2017 and does not duplicate any symbolic name in the current charmap (if any), or any other  
 2018 symbolic name defined in this collation definition. A <collating-symbol> defined via this  
 2019 keyword is only defined within the LC\_COLLATE category. More than one <collating-  
 2020 symbol> may be defined with one "collating-symbol" keyword, and symbolic ellipses may  
 2021 be used.

2022

2023 Example:  
 2024 collating-symbol <CAPITAL>  
 2025 collating-symbol <HIGH>

2026

#### 2027 4.4.7 "symbol-equivalence" keyword

2028

2029 This keyword is used to define symbols for use in collation sequence statements; and  
 2030 assign the same weight as another defined symbol. The syntax is

2031

2032 "symbol-equivalence %s %s\n", <collating-symbol-1>, <collating-symbol-2>

2033

The <collating-symbol-1> and <collating-symbol-2> are symbolic names, enclosed between angle brackets (< and >). <collating-symbol-1> does not duplicate any symbolic name in the current charmap (if any), or any other symbolic name defined in this collation definition. <collating-symbol-2> is defined elsewhere in the LC\_COLLATE category as a collating-symbol. The use of <collating-symbol-2> is equivalent to using the <collating-symbol-1> in the LC\_COLLATE category. A <collating-symbol-1> defined via this keyword is only defined within the LC\_COLLATE category.

Example  
 collating-symbol <CAP>  
 symbol-equivalence <CAPITAL> <CAP>

#### 4.4.8 "order\_start" keyword

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

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

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

and

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

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 are separated by semicolons (;). Each operand consists 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 issues a warning message. The following directives are supported:

- 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.
- 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.
- 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 collates first. If both strings contain a non-"IGNORE"d character in the same relative position, the collating values assigned to the elements determine the ordering. In case of equality, subsequent non-IGNOREd characters are considered in the same manner.

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

2087 Examples:  
 2088 order\_start forward;backward  
 2089 order\_start <CYRILLIC>;forward;forward  
 2090

2091 If no operands are specified, a single forward operand is assumed.  
 2092  
 2093

#### 2094 4.4.9 "order\_end" keyword

2095

2096 The collating order entries are terminated with an "order\_end" keyword.  
 2097

#### 2098 4.4.10 "reorder-after" keyword

2099

2100 The "reorder-after" keyword is used to specify a modification to a copied collation  
 2101 specification of an existing FDCC-set. There can be more than one "reorder-after"  
 2102 statement in a collating specification. The syntax is:  
 2103

```
2104 "reorder-after %s\n",<collating-symbol>
```

2105

2106 The <collating-symbol> operand is a symbolic name, enclosed between angle brackets,  
 2107 and is present in the source FDCC-set copied via the "copy" keyword.  
 2108 The "reorder-after" statement is followed by one or more collation statements as described  
 2109 in the "Collating Order" clause (4.4.5), with the exception that the ellipsis symbol (...) is  
 2110 not used.  
 2111

2112 Each collation statement reassigns character collation values and collation weights to  
 2113 collating elements existing in the copied collation specification, by removing the collating  
 2114 statement from the copied specification, and inserting the collating element in the collating  
 2115 sequence with the new collation weights after the preceding collating element of the  
 2116 "reorder-after" specification, the first collating element in the collation sequence being the  
 2117 <collating-symbol> specified in the "reorder-after" statement.  
 2118

2119 A "reorder-after" specification is terminated by another "reorder-after" specification or the  
 2120 "reorder-end" statement.  
 2121

##### 2122 4.4.10.1 Example of "reorder-after"

```
2123
2124   reorder-after <y8>
2125     <U:>      <Y>;<U:>;<CAPITAL>
2126     <u:>      <Y>;<U:>;<SMALL>
2127   reorder-after <z8>
2128     <AE>      <AE>;<NONE>;<CAPITAL>
2129     <ae>      <AE>;<NONE>;<SMALL>
2130     <A:>      <AE>;<DIAERESIS>;<CAPITAL>
2131     <a:>      <AE>;<DIAERESIS>;<SMALL>
2132     <O:/>    <O/>;<NONE>;<CAPITAL>
2133     <o:/>    <O/>;<NONE>;<SMALL>
2134     <AA>      <AA>;<NONE>;<CAPITAL>
2135     <aa>      <AA>;<NONE>;<SMALL>
2136   reorder-end
2137
```

2138 The example is interpreted as follows (using the "i18nrep" repertoiremap):  
 2139

1. The collating element <U:> is removed from the copied collating sequence and inserted after <y8> in the collating sequence with the new weights. The collating element <u:> is removed from the copied collating sequence and inserted in the resulting collation sequence after <U:> with the new weights. <y8> is used to indicate the last entry of the <y> letters.

2145    2. The second "reorder-after" statement terminates the first list of reordering collation identifier entries, and  
 2146    initiates a second list, rearranging the order and weights for the <AE>, <ae>, <A:>, <a:>, <O/>, and <o/>  
 2147    collating elements after the <z8> collating symbol in the copied specification. <z8> is used to indicate the  
 2148    last entry of the <z> letters.

2149  
 2150    3. The "reorder-end" statement terminates the second list of reordering entries.

2151  
 2152    4. Thus for the original sequence

2153    ... ( U u Ü ü ) V v W w X x Y y Z z

2155    this example reordering gives

2157    ... U u V v W w X x ( Y y Ü ü ) Z z ( Ä æ Å ä ) Ø ø Å å

2159    where the parenthesis indicate ordering with the same weight on the first level for multiple upper/lowercase  
 2160 pairs.

#### 2162    4.4.11 "reorder-end" keyword

2164    The "reorder-end" keyword specifies the end of a list of collating statements, initiated by  
 2165 the "reorder-after" keyword.

#### 2167    4.4.12 "reorder-section-after" keyword

2170    The "reorder-section-after" keyword is used to specify a modification to a copied collation  
 2171 specification of an existing FDCC-set. The "reorder-section-after" statement is followed by  
 2172 one or more statements consisting of section reordering statements.

##### 2174    4.4.12.1 section reordering statements

2176    The section reordering statements rearranges the set of collating entries and changes  
 2177 sorting rules for the set of collating entries identified by a section symbol in a preceding  
 2178 "order\_start" statement. Each section reorder statement has the syntax:

2180    "%s %s;...%s\n", <section-symbol>, <sort-rule>, <sort-rule> ...

2182    The <section-symbol> identifies the set of collating entries, and is defined via a "section-  
 2183 symbol" keyword.

2185    The <sort-rule>s are as described for the "order\_start" keyword. Specified <sort-rule>s  
 2186 replace the specification for the ordering of the section given on the "order\_start"  
 2187 statement identified by the <section-symbol>. The <sort-rule>s are optional, and <sort-  
 2188 rule>s not to be changed may be given by empty specifications.

2189    Note: The <sort-rule> capability is an extension over ISO/IEC 14651 functionality.

2192    The order of the section reordering statements rearranges the assignment of collation  
 2193 entries for the sets of collation entries identified by the <section-symbols> to the order  
 2194 that the <section-symbols> occur after the "reorder-section-after" statement.

2196    The section reordering statements are terminated by a "reorder-section-end" statement.

##### 2198    4.4.12.2 Example of section reordering

2199    copy "i18n"

```

2201 reorder-section-after <DIGITS>
2202 <ARABIC>
2203 <LATIN> forward;backward;forward;forward,position
2204 reorder-section-end
2205

```

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

#### 4.4.13 "reorder-section-end" keyword

The "reorder-section-end" keyword specifies the end of a list of section symbols, initiated by the "reorder-section-after" keyword.

#### 4.4.14 "i18n" LC\_COLLATE category

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

```

LC_COLLATE
% This is the ISO/IEC TR 14652 i18n fdcc-set definition for
% the LC_COLLATE category.
%
% equivalences
symbol-equivalence <NONE> <BLANK>
symbol-equivalence <CAPITAL> <CAP>
symbol-equivalence <SMALL> <MIN>
symbol-equivalence <CAPITAL-SMALL> <COMPATCAP>
symbol-equivalence <SMALL-CAPITAL> <COMPAT>
symbol-equivalence <MACRON> <MACRO>
symbol-equivalence <STROKE> <OBLIK>
symbol-equivalence <ACUTE> <AIGUT>
symbol-equivalence <CIRCUMFLEX> <CIRCF>
symbol-equivalence <RING> <CRCLE>
symbol-equivalence <DIAERESIS> <TREMA>
symbol-equivalence <DOT> <POINT>
symbol-equivalence <CEDILLA> <CEDIL>
symbol-equivalence <OGONEK> <OGONK>
symbol-equivalence <HOOK> <CROOK>
symbol-equivalence <HORN> <HORNU>
symbol-equivalence <DOT-BELOW> <POINS>

order_start forward;forward;forward;forward,position

% Copy the template from ISO/IEC 14651
copy "ISO14651_2000_TABLE1.txt"

order_end

END LC_COLLATE

```

#### 4.5 LC\_MONETARY

The LC\_MONETARY category defines the rules and symbols that are 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, are used to indicate that the value is unspecified, and then no default is implied. The following keywords are 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 is specified.

**valid\_from**

One or more strings separated by semicolons, representing a Gregorian date in the form "YYYYMMDD" according to ISO 8601, specifying the beginning date (inclusive from the beginning of day local time) of the validity of a currency.

The position of the string 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 an implementation-defined beginning of time. This keyword is optional.

**valid\_to**

One or more strings separated by semicolons, representing a Gregorian date in the form "YYYYMMDD" according to ISO 8601, specifying the end date (inclusive to the end of day local time) of the validity of a currency. If not specified, it is taken to be an implementation-defined end of time. This keyword is optional.

**conversion\_rate**

one or more pairs of integers separated by a <semicolon> specifying the fixed conversion rate between the current currency (determined by the parameter number) and the first currency that is valid, determined by a date provided by the application. If the currency is not the first valid currency for the period in question, the first integer is for multiplying the first valid currency, and the second for dividing this result to get the amount in the current currency. The currency to be the current currency is selected by the application from the date applicable and the currency number (first, second, third etc valid currency at that date); and whether domestic or international formatting is used is also determined by the application. Each pair of integers are separated by a <slash>. The default value is "1/100". This keyword is optional.

Note: The two integers are used instead of a floating point value, to be able to cater for legal requirements on Euro conversion where a multiplication and division is prescribed, instead of just one floating point multiplication.

**currency\_symbol**

One or more strings separated by semicolons that are used as the local currency symbol.

**mon\_decimal\_point**

The operand is a string containing the symbol that is 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 is specified, unless the "copy" keyword is used.

**mon\_thousands\_sep**

The operand is a string containing the symbol that is used as a separator for groups of digits to the left of the decimal

2317		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 is specified, unless the "copy" keyword is used.
2318		
2319		
2320		
2321		
2322	<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) is repeatedly used for the remainder of the digits. If the last integer is -1, then no further grouping is performed. The keyword is specified, unless the "copy" keyword is used.
2323		
2324		
2325		
2326		
2327		
2328		
2329		
2330		
2331		
2332		
2333	<b>positive_sign</b>	A string that is used to indicate a nonnegative-valued formatted monetary quantity. The keyword is specified, unless the "copy" keyword is used.
2334		
2335		
2336	<b>negative_sign</b>	A string that is used to indicate a negative-valued formatted monetary quantity. The keyword is specified, unless the "copy" keyword is used.
2337		
2338		
2339	<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 is specified, unless the "copy" keyword is used.
2340		
2341		
2342		
2343		
2344	<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 is specified, unless the "copy" keyword is used.
2345		
2346		
2347		
2348		
2349	<b>p_sep_by_space</b>	One or more integers separated by semicolons, set to 0 if no space separates the "currency_symbol" from the value for a nonnegative formatted monetary quantity, set to 1 if a space separates the symbol from the value, and set to 2 if a space separates the symbol and the sign string, if adjacent. The keyword is specified, unless the "copy" keyword is used.
2350		
2351		
2352		
2353		
2354		
2355	<b>n_cs_precedes</b>	One or more integers separated by semicolons, set to 1 if the "currency_symbol" precedes the value for a negative formatted monetary quantity, and set to 0 if the symbol succeeds the value. The keyword is specified, unless the "copy" keyword is used.
2356		
2357		
2358		
2359		
2360	<b>n_sep_by_space</b>	One or more integers separated by semicolons, set to 0 if no space separates the "currency_symbol" from the value for a negative formatted monetary quantity, set to 1 if a space separates the symbol from the value, and set to 2 if a space separates the symbol and the sign string, if adjacent. The keyword is specified, unless the "copy" keyword is used.
2361		
2362		
2363		
2364		
2365		
2366	<b>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 monetary quantity using the
2367		
2368		

2369		"currency_symbol". The following integer values are defined:
2370		
2371	0	Parentheses enclose the quantity and the "currency_symbol".
2372	1	The sign string precedes the quantity and the "currency_symbol".
2373	2	The sign string succeeds the quantity and the "currency_symbol".
2374	3	The sign string immediately precedes the "currency_symbol".
2375	4	The sign string immediately succeeds the "currency_symbol".
2376		The keyword is specified, unless the "copy" keyword is used.
2377		
2378		
2379		
2380		
2381		
2382		
2383	<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 are defined:
2384		
2385	0	Parentheses enclose the quantity and the "currency_symbol".
2386	1	The sign string precedes the quantity and the "currency_symbol".
2387	2	The sign string succeeds the quantity and the "currency_symbol".
2388	3	The sign string immediately precedes the "currency_symbol".
2389	4	The sign string immediately succeeds the "currency_symbol".
2390		The keyword is specified, unless the "copy" keyword is used.
2391		
2392		
2393		
2394		
2395		
2396		
2397		
2398		
2399		
2400	<b>int_curr_symbol</b>	One or more strings separated by semicolons that are used as the international currency symbols. Each operand is a four character string, with the first three characters containing the alphabetic international currency symbol in accordance with those specified in ISO 4217, <i>Codes for the representation of currencies and funds</i> . The fourth character is the character used to separate the international currency symbol from the monetary quantity. The keyword is specified, unless the "copy" keyword is used.
2401		
2402		
2403		
2404		
2405		
2406		
2407		
2408		
2409	<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 is specified, unless the "copy" keyword is used.
2410		
2411		
2412		
2413		
2414	<b>int_p_cs_precedes</b>	One or more integers separated by semicolons; set to 1 if the "int_curr_symbol" precedes the value for a nonnegative formatted monetary quantity, and set to 0 if the symbol succeeds the value. If not specified, the value of "p_cs_precedes" is taken.
2415		
2416		
2417		
2418		
2419	<b>int_p_sep_by_space</b>	One or more integers separated by semicolons; set to 0 if no space separates the "int_curr_symbol" from the value for a
2420		

2421		nonnegative formatted monetary quantity, set to 1 if a space separates the symbol from the value, and set to 2 if a space separates the symbol and the sign string, if adjacent. If not specified, the value of "p_sep_by_space" is taken.
2422		
2423		
2424		
2425	<b>int_n_cs_precedes</b>	One or more integers separated by semicolons; set to 1 if the "int_curr_symbol" precedes the value for a negative formatted monetary quantity, and set to 0 if the symbol succeeds the value. If not specified, the value of "n_cs_precedes" is taken.
2426		
2427		
2428		
2429		
2430	<b>int_n_sep_by_space</b>	One or more integers separated by semicolons; set to 0 if no space separates the "int_curr_symbol" from the value for a negative formatted monetary quantity, set to 1 if a space separates the symbol from the value, and set to 2 if a space separates the symbol and the sign string, if adjacent. If not specified, the value of "n_sep_by_space" is taken.
2431		
2432		
2433		
2434		
2435		
2436	<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 monetary quantity using the "int_curr_symbol". The following integer values are defined:
2437		
2438		
2439		
2440		
2441	0	Parentheses enclose the quantity and the "int_curr_symbol".
2442	1	The sign string precedes the quantity and the "int_curr_symbol".
2443	2	The sign string succeeds the quantity and the "int_curr_symbol".
2444	3	The sign string immediately precedes the "int_curr_symbol".
2445	4	The sign string immediately succeeds the "int_curr_symbol".
2446		If no "int_p_sign_posn" is present the value of the "p_sign_posn" is taken.
2447		
2448		
2449		
2450		
2451		
2452		
2453		
2454	<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 monetary quantity using the "int_curr_symbol". The following integer values are defined:
2455		
2456		
2457		
2458		
2459	0	Parentheses enclose the quantity and the "int_curr_symbol".
2460	1	The sign string precedes the quantity and the "int_curr_symbol".
2461	2	The sign string succeeds the quantity and the "int_curr_symbol".
2462	3	The sign string immediately precedes the "int_curr_symbol".
2463	4	The sign string immediately succeeds the "int_curr_symbol".
2464		If no "int_n_sign_posn" is present the value of the "n_sign_posn" is taken.
2465		
2466		
2467		
2468		
2469		
2470		
2471		
2472	The "i18n" FDCC-set is defined as follows for the LC_MONETARY category.	

```

2473
2474     LC_MONETARY
2475     % This is the 14652 i18n fdcc-set definition for
2476     % the LC_MONETARY category.
2477     %
2478     int_curr_symbol      ""
2479     currency_symbol      ""
2480     mon_decimal_point    "<U002C>" 
2481     mon_thousands_sep     ""
2482     mon_grouping         -1
2483     positive_sign        ""
2484     negative_sign        "<U002E>" 
2485     int_frac_digits       -1
2486     frac_digits          -1
2487     p_cs_precedes        -1
2488     p_sep_by_space        -1
2489     n_cs_precedes        -1
2490     n_sep_by_space        -1
2491     p_sign_posn          -1
2492     n_sign_posn          -1
2493     %
2494 END LC_MONETARY
2495
2496

```

## 4.6 LC\_NUMERIC

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

<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 is specified.
<b>decimal_point</b>	The operand is a string containing the symbol that is used as the decimal delimiter in numeric, nonmonetary formatted quantities. This keyword cannot be omitted and cannot be set to the empty string. In contexts where other standards limit the decimal point to a single byte, the result of specifying a multibyte operand is unspecified.
<b>thousands_sep</b>	The operand is a string containing the symbol that is used as a separator for groups of digits to the left of the decimal delimiter in numeric, nonmonetary formatted monetary quantities. In contexts where other standards limit the "thousands_sep" to a single byte, the result of specifying a multibyte operand is unspecified.
<b>grouping</b>	Define the size of each group of digits in formatted non-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) is repeatedly used for the remainder of the digits. If the last integer is -1, then no further grouping is performed.

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

```

2531
2532     LC_NUMERIC
2533     % This is the 14652 i18n fdcc-set definition for
2534     % the LC_NUMERIC category.
2535     %
2536     decimal_point    "<U002C>"
2537     thousands_sep   ""
2538     grouping        -1
2539     %
2540 END LC_NUMERIC
2541

```

## 4.7 LC\_TIME

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

2548 <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 is specified.
2551 <b>abday</b>	Define the abbreviated weekday names for calendar systems with weeks of constant length, to be referenced by the %a field descriptor. The length of the week and a gregorian date for the first weekday is defined by the "week" keyword. The operand consists of semicolon-separated strings. The first string is the abbreviated name of the day corresponding to the first day of the week (default Sunday), the second the abbreviated name of the day corresponding to the second day of the week (default Monday), and so on.
2559 <b>day</b>	Define the full weekday names for calendar systems with weeks of constant length, to be referenced by the %A field descriptor. The length of the week and a gregorian date for the first weekday is defined by the "week" keyword. The operand consists of semicolon-separated strings. The first string is the full name of the day corresponding to the first day of the week (default Sunday), the second the full name of the day corresponding to the second day of the week (default Monday), and so on.
2567 <b>week</b>	Is used to define the number of days in a week, and which weekday is the first weekday (the first weekday has the value 1), and which week is to be considered the first in a year. The first operand is an integer specifying the number of days in the week. The second operand is an integer specifying the Gregorian date in the format YYYYMMDD, and it specifies a day that is a first weekday (all other first weekdays may then be calculated by adding or subtracting a whole multiplum of the number of days in the week as specified with the first operand). The third operand is an integer specifying the weekday number to be contained in the first week of the year. The third operand may also be understood as the number of days required in a week for it to be considered the first week of the year. If the keyword is not specified the values are taken as 7, 19971130 (a Sunday), and 7 (Saturday), respectively. ISO 8601 conforming applications should use the values 7, 19971201 (a Monday), and 4 (Thursday), respectively. This keyword is optional.
2583 <b>abmon</b>	Define the abbreviated month names, to be referenced by the %b field descriptor. The operand consists of twelve or thirteen semicolon-separated strings. The first string is the abbreviated name

2586	of the first month of the year (January), the second the abbreviated name of the second month, and so on.
2587	
2588	<b>mon</b> Define the full month names, to be referenced by the %B field descriptor. The operand consists of twelve or thirteen semicolon-separated strings. The first string is the full name of the first month of the year (January), the second the full name of the second month, and so on.
2589	
2590	
2591	
2592	
2593	<b>d_t_fmt</b> Define the appropriate date and time representation, to be referenced by the %c field descriptor. The operand consists of a string, and can contain any combination of characters and field descriptors. In addition, the string can contain escape sequences defined in Table 3.
2594	
2595	
2596	
2597	<b>d_fmt</b> Define the appropriate date representation, to be referenced by the %x field descriptor. The operand consists of a string, and can contain any combination of characters and field descriptors. In addition, the string can contain escape sequences defined in Table 3.
2598	
2599	
2600	
2601	<b>t_fmt</b> Define the appropriate time representation, to be referenced by the %X field descriptor. The operand consists of a string, and can contain any combination of characters and field descriptors. In addition, the string can contain escape sequences defined in Table 3.
2602	
2603	
2604	
2605	<b>am_pm</b> Define the appropriate representation of the ante meridiem and post meridiem strings, to be referenced by the %p field descriptor. The operand consists of two strings, separated by a semicolon. The first string represents the antemeridiem designation, the last string the postmeridiem designation. The keyword is optional. If unspecified, the %p field descriptor refers to the empty string.
2606	
2607	
2608	
2609	
2610	
2611	<b>t_fmt_ampm</b> Define the appropriate time representation in the 12-hour clock format with "am_pm", to be referenced by the %r field descriptor. The operand consists of a string and can contain any combination of characters and field descriptors. If the string is empty, the 12-hour format is not supported in the FDCC-set.
2612	
2613	
2614	
2615	
2616	

### The following keywords are all optional

2617	<b>era</b> Is used to define alternate Eras, corresponding to the %E field descriptor modifier. The format of the operand is unspecified, but supports the definition of the %EC and %Ey field descriptors, and may also define the "era_year" format (%EY).
2618	
2619	<b>era_year</b> Is used to define the format of the year in alternate Era format, corresponding to the %EY field descriptor.
2620	
2621	
2622	
2623	<b>era_d_t_fmt</b> Is used to define the format of the date and time in alternate Era notation, corresponding to the %Ec field descriptor.
2624	
2625	<b>era_d_fmt</b> Is used to define the format of the date in alternate Era notation, corresponding to the %Ex field descriptor.
2626	
2627	<b>era_t_fmt</b> Is used to define the format of the time in alternate Era notation, corresponding to the %EX field descriptor.
2628	
2629	
2630	
2631	<b>alt_digits</b> Is used to define alternate symbols for digits, corresponding to the %O field descriptor modifier. The operand consists of semicolon-separated strings. The first string is the alternate symbol corresponding with zero, the second string the symbol corresponding with one, and so on. Up to 100 alternate symbol strings can be specified. The %O modifier indicates that the string corresponding to the value specified via the field descriptor is used instead of the value.
2632	
2633	
2634	
2635	
2636	
2637	

2638	<b>first_weekday</b>	Is used to define the first day to be displayed, for example in a calendar display utility. The operand is an integer specifying the day number (1 = first) according to the information specified with the "day" keyword. The keyword may be omitted, and then the value 1 is taken, corresponding to Sunday for a week beginning Sunday, or to Monday for a week beginning Monday.
2639		
2640		
2641		
2642		
2643		
2644	<b>first_workday</b>	Is used to define the first workday as an integer according to the day numbering specified with the "week" keyword.
2645		
2646	<b>cal_direction</b>	Is used to define the direction of the display of dates, for example in a calendar display utility. The operand is an integer, and the following values are defined:
2647		1 left-right from top
2648		2 top-down from left
2649		3 right-left from top
2650		The keyword may be omitted, and then the value 1 is taken.
2651		
2652	<b>timezone</b>	Is used to define one or more timezones, each defined by a string, and the strings separated by a <semicolon>. In the following the characters <, >, [ and ] are used as metacharacters. Only characters with a visible glyph from the portable character set may be used, except in the <std> and <dst> fields. The syntax of a string is:
2653		
2654		
2655		
2656		
2657		
2658		
2659		<std><offset><dst>[<offset>][,<rule>[,<rule>...]];
2660		
2661		where
2662		
2663		<std> and <dst> Indicates no less than three, nor more than 10 characters that are the designation for the standard <std>, or Daylight Savings Time or summer time <dst> zone. Only <std> is required; if <dst> is missing, then Daylight Savings Time or summer time does not apply in this category. Upper- and lowercase letters are explicitly allowed. Any characters except a leading colon <:> or digits, the comma <,>, the minus <->, the plus <+>, and the null character are permitted to appear in these fields, but their meaning is unspecified.
2664		
2665		
2666		
2667		
2668		
2669		
2670		
2671		
2672		
2673		
2674		
2675	<b>&lt;offset&gt;</b>	Indicates the value one must add to the local time to arrive at the Coordinated Universal Time. The <offset> has the form:
2676		
2677		
2678		
2679		hh[:mm[:ss]]
2680		
2681		The minutes (mm) and seconds (ss) are optional. The hour (hh) is required and may be a single digit. The <offset> following <std> is required. If no <offset> follows <dst>, summer time is assumed to be one hour ahead of standard time. One or more digits may be used; the value is always interpreted as a decimal number. The hour is between zero and 24, and the minutes (and seconds) - if present - is
2682		
2683		
2684		
2685		
2686		
2687		
2688		
2689		

2690  
2691  
2692  
2693  
2694 <rule>  
2695  
2696  
2697

between zero and 59. If preceded by a "-", the time zone is east of the Prime Meridian; otherwise it is west of (which may be indicated by an optional preceding "+").

A specification for Daylight Savings Time changes that indicates when to change to and back from summer time. The <rule> has the form:

<date>[/<time>/<year>],<date>[/<time>/<year>]

where the first <date> describes when the change from standard time to summer time occurs, and the second <date> describes when the change back happens. Each <time> field describes when, in current local time, the change to the other time is made. The first <year> field defines the beginning of the validity of this rule, and the second <year> field defines the end of the validity of the rule. A number of rules may be given.

The format of <date> is one of the following:

J<n> The Julian day <n> (1 <= n <= 365) Leap years are not counted. That is, in all years - including leap years - February 28 is day 59 and March 1 is day 60. It is impossible to explicitly refer to the occasional February 29.  
<n> The zero-based Julian day (0 <= n <= 365). Leap years are counted and it is possible to refer to February 29.

M<m>.<n>.<d> the <d>th day (0 <= d <= 7) of week <n> of month <m> (1 <= n <= 5, 1 <= m <= 12, where week 5 means "the last <d> day in month <m>" which may occur in either the fourth or fifth week). Week 1 is the first week in which the <d>th day occurs. Day zero and day seven is Sunday.

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

2742 The <year> has the format YYYY.  
2743  
2744 NOTE: This way of specifying the timezone is compatible with the  
2745 format for the environment variable TZ described in Section 8.1.1 of  
2746 POSIX.1.

#### 4.7.1 Date Field Descriptors

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

**Table 3: Escape sequences for the date field**

%a	FDCC-set's abbreviated weekday name.
%A	FDCC-set's full weekday name.
%b	FDCC-set's abbreviated month name.
%B	FDCC-set's full month name.
%c	FDCC-set's appropriate date and time representation.
%C	Century (a year divided by 100 and truncated to integer) as decimal number (00-99).
%d	Day of the month as a decimal number (01-31).
%D	Date in the format mm/dd/yy.
%e	Day of the month as a decimal number (1-31 in at two-digit field with leading <space> fill).
%F	The date in the format YYYY-MM-DD (ISO 8601 format).
%g	Week-based year within century, as a decimal number (00-99).
%G	Week-based year with century, as a decimal number (for example 1997).
%h	A synonym for %b.
%H	Hour (24-hour clock), as a decimal number (00-23).
%I	Hour (12-hour clock), as a decimal number (01-12).
%j	Day of the year, as a decimal number (001-366).
%m	Month, as a decimal number (01-13).
%M	Minute, as a decimal number (00-59).
%n	A <newline> character.
%p	FDCC-set's equivalent of either AM or PM.
%r	12-hour clock time (01-12), using the AM/PM notation.
%R	24-hour clock time, in the format "%H:%M".
%S	Seconds, as a decimal number (00-61).
%t	A <tab> character.
%T	24-hour clock time, in the format HH:MM:SS.
%u	Weekday, as a decimal number (1(Monday)-7).
%U	Week number of the year (Sunday as the first day of the week) as a decimal number (00-53). All days in a new year preceding the first Sunday are considered to be in week 0.
%v	Week number of the year, as a decimal number with two digits including a possible leading zero, according to "week" keyword.
%V	Week of the year (Monday as the first day of the week), as a decimal number (01-53). The method for determining the week number is as specified by ISO 8601.

2794	%w	Weekday, as a decimal number (0(Sunday)-6).
2795	%W	Week number of the year (Monday as the first day of the week), as a decimal number (00-53). All days in a new year preceding the first Monday are considered to be in week 0.
2796		
2797		
2798	%x	FDCC-set's appropriate date representation.
2799	%X	FDCC-set's appropriate time representation.
2800	%y	Year within century (00-99).
2801	%Y	Year with century, as a decimal number.
2802	%z	The offset from UTC in the ISO 8601 format "-0430" (meaning 4 hours 30 minutes behind UTC, west of Greenwich), or by no characters if no time zone is determinable.
2803		
2804		
2805	%Z	Time-zone name, or no characters if no time zone is determinable.
2806	%%	A <percent-sign> character.
2807		
2808	NOTE: %g, %G and %V give values according to the ISO 8601 week-based year. In this system, weeks begin on a Monday and week 1 of the year is the week that includes 4th January, which is also the week that includes the first Thursday of the year, and is also the first week that contains at least four days in the year. If the first Monday of the year is the 2nd, 3rd or 4th, the preceding days are part of the last week of the preceding year; thus, for Saturday 2nd January 1999, %G is replaced by 1998 and %V is replaced by 53. If the 29th, 30th or 31st January is a Monday, it and any following days are part of week 1 of the following year. Thus, for Tuesday 30th December 1997, %G is replaced by 1998 and %V is replaced by 1.	
2809		
2810		
2811		
2812		
2813		
2814		
2815		
2816		
2817		

#### 4.7.2 Modified Field Descriptors

2818	%Ec	FDCC-set's alternate date and time representation.
2819	%EC	The name of the base year (period) in the FDCC-set's alternate representation.
2820		
2821		
2822		
2823		
2824		
2825		
2826	%Ex	FDCC-set's alternate date representation.
2827	%EX	FDCC-set's alternate time representation.
2828		
2829	%Ey	Offset from %EC (year only) in the FDCC-set's alternate representation.
2830	%EY	Full alternate year representation.
2831		
2832	%Od	Day of month using the FDCC-set's alternate numeric symbols.
2833	%Oe	Day of month using the FDCC-set's alternate numeric symbols.
2834	%Of	Weekday as a decimal number according to alt_day (1 is first day).
2835	%OH	Hour (24-hour clock) using the FDCC-set's alternate numeric symbols.
2836	%OI	Hour (12-hour clock) using the FDCC-set's alternate numeric symbols.
2837	%Om	Month using the FDCC-set's alternate numeric symbols.
2838	%OM	Minutes using the FDCC-set's alternate numeric symbols.
2839	%OS	Seconds using the FDCC-set's alternate numeric symbols.
2840	%Ou	Weekday as a number in the alternate representation of the FDCC-set (Monday=1).
2841	%OU	Week number of the year (Sunday as the first day of the week) using the FDCC-set's alternate numeric symbols.
2842		
2843	%OV	Week number of the year (Monday as the first day of the week, ISO 8601 rules)
2844		
2845		

2846 using the alternate numeric symbols of the FDCC-set.  
 2847 %Ow Weekday as number in the FDCC-set's alternate representation  
 2848 (Sunday=0).  
 2849 %OW Week number of the year (Monday as the first day of the week) using the  
 2850 FDCC-set's alternate numeric symbols.  
 2851 %Oy Year (offset from %C) in alternate representation.

#### 2853 4.7.3 "i18n" LC\_TIME category

2855 The "i18n" LC\_TIME category is (following ISO 8601):

```

2857 LC_TIME
2858   % This is the ISO/IEC TR 14652 "i18n" definition for
2859   % the LC_TIME category.
2860   %
2861   % Weekday and week numbering according to ISO 8601
2862   abday  "<U0031>" ; "<U0032>" ; "<U0033>" ; "<U0034>" ; /
2863   " <U0035>" ; "<U0036>" ; "<U0037>" ;
2864   day    "<U0031>" ; "<U0032>" ; "<U0033>" ; "<U0034>" ; /
2865   " <U0035>" ; "<U0036>" ; "<U0037>" ;
2866   week   7;19971201;4
2867   abmon  "<U0030><U0031>" ; "<U0030><U0032>" ; "<U0030><U0033>" ; /
2868   " <U0030><U0034>" ; "<U0030><U0035>" ; "<U0030><U0036>" ; /
2869   " <U0030><U0037>" ; "<U0030><U0038>" ; "<U0030><U0039>" ; /
2870   mon    "<U0031><U0030>" ; "<U0031><U0031>" ; "<U0031><U0032>" ;
2871   " <U0030><U0031>" ; "<U0030><U0032>" ; "<U0030><U0033>" ; /
2872   " <U0030><U0034>" ; "<U0030><U0035>" ; "<U0030><U0036>" ; /
2873   " <U0030><U0037>" ; "<U0030><U0038>" ; "<U0030><U0039>" ; /
2874   " <U0031><U0030>" ; "<U0031><U0031>" ; "<U0031><U0032>" ;
2875   am_pm  " " ;
2876   %
2877   % Date formats following ISO 8601
2878   % Appropriate date and time representation (%c)
2879   %   "%F %T"
2880   d_t_fmt "<U0025><U0046><U0020><U0025><U0054>" ;
2881   %
2882   % Appropriate date representation (%x)      "%F"
2883   d_fmt  "<U0025><U0046>" ;
2884   %
2885   % Appropriate time representation (%X)      "%T"
2886   t_fmt  "<U0025><U0054>" ;
2887   t_fmt_ampm " " ;
2888   %
2889   END LC_TIME
  
```

#### 2890 4.8 LC\_MESSAGES

2891 The LC\_MESSAGES category defines the format and values for affirmative and negative  
 2892 responses. The operands are strings or extended regular expressions to specify which  
 2893 response strings that should be considered matches; see ISO/IEC 9945-2:1993 clause 2.8.4  
 2894 for a definition of extended regular expressions. The following keywords are defined:  
 2895

2896 **copy** Specify the name of an existing FDCC-set to be used as the source for the  
 2897 definition of this category. If this keyword is specified, no other keyword  
 2898 is specified.  
 2899 **yesexpr** The operand consists of an extended regular expression that describes the  
 2900 acceptable affirmative response to a question expecting an affirmative or  
 2901 negative response.  
 2902 **noexpr** The operand consists of an extended regular expression that describes the  
 2903 acceptable negative response to a question expecting an affirmative or  
 2904 negative response.  
 2905

2906 The "i18n" LC\_MESSAGES category is:

2907

2908

```

2909 LC_MESSAGES
2910 % This is the ISO/IEC 14652 "i18n" definition for
2911 % the LC_MESSAGES category.
2912 %
2913 yesexpr "<U005B><U002B><U0031><U005D>"
2914 noexpr " <U005B><U002D><U0030><U005D>""
2915 END LC_MESSAGES
2916

```

2917 Note: This uses regular expression syntax with brackets [...] to for example  
 2918 specify the both <+> and <1> is allowed as an affirmative answer.

## 2920 4.9 LC\_XLITERATE

2922 The LC\_XLITERATE category defines formats to transliterate strings, by transforming  
 2923 substrings in the source to substrings in the target string. The capabilities are limited to  
 2924 simple transliteration based on substring substitution, while more advanced transliteration  
 2925 schemes, for example based on pattern matching, is either cumbersome to specify, or not  
 2926 addressed. The transliteration may for example be from the Cyrillic script to the Latin  
 2927 script.

2929 Transliteration of an incoming character string to a character string in a FDCC-set can be  
 2930 specified with the following transliteration keywords and transliteration statements.

2932 <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 is specified.
2933 <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.
2934 <b>default_missing</b>	defines a string of one or more characters to be put in the output string if no transliteration statement can be applied to a input <transliteration-source>. This keyword is optional.
2935 <b>translit_ignore</b>	defines a set of characters, separated by semicolons, that are to be ignored in the incoming character string, that is, each of the occurrences of such characters is treated as the empty string. The characters may use the notations defined in 4.3 for lists of characters. This keyword is optional.
2936 <b>redefine</b>	This keyword introduces a list of transliteration statements where each of the <transliteration_source> strings have been defined previously in the specification, and the new transliteration statements then replaces the old transliteration statements for the <transliteration_source> strings specified. This keyword is optional.

### 2955 4.9.1 Transliteration statements

2956 The syntax for a transliteration statement is:

```
"%s %s;%s;...;%s\n",<transliteration_source>,<transliteration_string>,...
```

2957 Each <transliteration\_source> consists of one or more characters (in any of the forms  
 2958 defined in 4.1.1). The <transliteration\_source> that is the longest in terms of number of  
 2959 characters that match the input string is the one selected for transliteration.

If a transliteration statement contains more than one <transliteration\_string>, the order that each <transliteration\_string> occurs in the transliteration statement defines the precedence order for choosing a particular <transliteration\_string> to substitute for the <transliteration\_source>. When a process makes use of a transliteration statement to transliterate text, and that transliteration statement contains more than one <transliteration\_string>, that process chooses the first <transliteration\_string>, in the defined precedence order, that satisfies the requirements of the transliteration.

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

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

#### 4.9.2 "include" keyword

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

The syntax of the "include" statement is:

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

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

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

#### 4.9.3 Example of use of transliteration

```
LC_XLITERATE
include "de_DE";"de_repmap"
default_missing <?>
translit_ignore <U3200>..<UFAFF>
<ae>      <a:>;<e*>;"<a><e>" ;<e>"
<s>        <s*>;<s=>
"<K><O>"   <KO>
END LC_XLITERATE
```

The "LC\_XLITERATE" statement introduces the transliteration category.

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

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

3022 The "translit\_ignore" keyword specifies that a set of Ideographic characters, Hangul, East Asian symbols and the  
 3023 private use area etc. (the range <U3200>..<UFAFF>) is ignored for the transliteration.  
 3024  
 3025 The next 3 lines are transliteration statements.  
 3026  
 3027 The first transliteration statement defines a number of transliterations for the LATIN LETTER AE, including into  
 3028 LATIN LETTER A WITH DIAERESIS, GREEK LETTER EPSILON, the two Latin letters A and E, and finally  
 3029 the LATIN LETTER E.  
 3030  
 3031 The second transliteration statement defines transliteration of the LATIN LETTER S into GREEK LETTER  
 3032 SIGMA, and CYRILLIC LETTER ES.  
 3033  
 3034 The third transliteration statement transliterates the two Latin letters K and O into the Japanese Hiragana character  
 3035 KO.  
 3036  
 3037 The transliteration category is terminated via the "END LC\_XLITERATE" statement in the above example.  
 3038  
 3039 There is no "i18n" entry for the LC\_XLITERATE category

## 4.10 LC\_NAME

3043 The LC\_NAME category defines formats to be used in addressing a person, e.g. in a  
 3044 postal address or in a letter. The following keywords are defined:  
 3045  
 3046 **copy** Specify the name of an existing FDCC-set to be used as the source for the  
 3047 definition of this category. If this keyword is specified, no other keyword is  
 3048 specified.  
 3049 **name\_fmt** Define the appropriate representation of a person's name and title. The  
 3050 operand consists of a string, and can contain any combination of characters  
 3051 and field descriptors. In addition, the string can contain escape sequences  
 3052 defined below.  
 3053 **name\_gen** The operand is a string defining a salutation valid for all persons.  
 3054 **name\_miss** The operand is a string defining a salutation valid for unmarried females.  
 3055 **name\_mr** The operand is a string defining a salutation valid for males.  
 3056 **name\_mrs** The operand is a string defining a salutation valid for married females.  
 3057 **name\_ms** The operand is a string defining a salutation valid for all females.

3059       NOTE: There are a number of variations for addressing a person among the  
 3060 cultures. Middle names are not used in many countries and even the family name  
 3061 is not used in some countries. In other countries there is extensive use of one or  
 3062 more middle names and corresponding initials. The specification below should be  
 3063 regarded as a starting point for this problem.

3065 The LC\_NAME category defines the interpretation of a number of escape sequences. The  
 3066 escape sequences are also available in the definitions with the following LC\_NAME  
 3067 keywords: "name\_fmt".

3069 Escape sequences for the "name\_fmt" keyword:

3071       %f Family names.  
 3072       %F Family names in uppercase.  
 3073       %g First given name.  
 3074       %G First given initial.  
 3075       %l First given name with latin letters.  
 3076       %o Other shorter name, eg. "Bill".

3077 %m Middle names.  
 3078 %M Middle initials.  
 3079 %p Profession.  
 3080 %s Salutation, such as "Doctor"  
 3081 %S Abbreviated salutation, such as "Mr." or "Dr."  
 3082 %d Salutation, using the FDCC-sets conventions, with 1 for the name\_gen, 2 for  
 3083 name\_mr, 3 for name\_mrs, 4 for name\_miss, 5 for name\_ms.  
 3084 %t If the preceding escape sequence resulted in an empty string, then the empty string,  
 3085 else a <space>.

3086  
 3087 Each escape sequence may have an <R> after the <%> to specify that the information is  
 3088 taken from a Romanized version string of the entity.

3089  
 3090 The "i18n" LC\_NAME category is:  
 3091

```
3092 LC_NAME
3093 % This is the ISO/IEC TR 14652 "i18n" definition for
3094 % the LC_NAME category.
3095 %
3096 name_fmt   "<U0025><U0070><U0025><U0074><U0025><U0067><U0025><U0074>/
3097 <U0025><U006D><U0025><U0074><U0025><U0066>"
3098 END LC_NAME
3099
```

## 4.11 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 related to geography. All keywords are optional. The following keywords are recognized:

3100 <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 is specified.
3101 <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 consists of a string, and can contain any combination of characters and field descriptors. In addition, the string can contain escape sequences defined below.
3102 <b>country_name</b>	The operand is a string with the name of the country in the language of the FDCC-set.
3103 <b>country_post</b>	The operand is a string with the abbreviation of the country, used for postal addresses, for example by CEPT-MAILCODE.
3104 <b>lang_name</b>	The operand is a string with the name of the language in the language of the FDCC-set.
3105 <b>lang_ab2</b>	The operand is a string with the two-letter abbreviation of the language, according to ISO 639.
3106 <b>lang_ab3_term</b>	The operand is a string with the three-letter abbreviation of the language for terminology use, according to ISO 639-2.
3107 <b>lang_ab3_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_ab3_term" keyword is taken.

Note: The "lang\_ab3\_term" and "lang\_ab3\_lib" keywords will in most cases contain the same value, but they may differ, e.g. the value for the German language is "deu" and "ger" respectively.

3132 The LC\_ADDRESS category defines the interpretation of a number of escape sequences.  
 3133 The escape sequences are also available in the definitions with the following  
 3134 LC\_ADDRESS keywords: "postal\_fmt".  
 3135  
 3136 Escape sequences for the "postal\_fmt" keyword:  
 3137  
 3138 %a C/O address.  
 3139 %f Firm name.  
 3140 %d Department name.  
 3141 %b Building name.  
 3142 %s Street or block (eg. Japanese) name.  
 3143 %h House number or designation.  
 3144 %N If any graphical characters have been specified then an end of line is  
 3145 made.  
 3146 %t If the preceding escape sequence resulted in an empty string, then the  
 3147 empty string, else a <space>.  
 3148 %r Room number, door designation.  
 3149 %e Floor number.  
 3150 %C Country designation, from the <country\_post> keyword.  
 3151 %l Local township  
 3152 %z Zip number, postal code.  
 3153 %T Town, city.  
 3154 %S State, province, or prefecture.  
 3155 %c Country.

3156  
 3157 Each escape sequence may have an <R> after the <%> to specify that the information is  
 3158 taken from a Romanized version string of the entity.  
 3159

3160 NOTE: There are a number of variations for specifying a location among the cultures.  
 3161 Some of the information, like the middle names, or even the family name, is not used  
 3162 in some cultures. The specification here should be regarded as a starting point for this  
 3163 problem.  
 3164

3165 The "i18n" LC\_ADDRESS category is:  
 3166  
 3167  
 3168 LC\_ADDRESS  
 3169 % This is the ISO/IEC TR 14652 "i18n" definition for  
 3170 % the LC\_ADDRESS category.  
 3171 %  
 3172 postal\_fmt "<U0025><U0061><U0025><U004E><U0025><U0066><U0025><U004E>/  
 3173 <U0025><U0064><U0025><U004E><U0025><U0062><U0025><U004E><U0025><U0073>/  
 3174 <U0020><U0025><U0068><U0020><U0025><U0065><U0020><U0025><U0072><U0025>/  
 3175 <U004E><U0025><U0043><U002D><U0025><U007A><U0020><U0025><U0054><U0025>/  
 3176 <U004E><U0025><U0063><U0025><U004E>"  
 3177 END LC\_ADDRESS  
 3178

## 3179 4.12 LC\_TELEPHONE

3181 The LC\_TELEPHONE category defines formats to be used with telephone services. All  
 3182 keywords are optional. The following keywords are defined:  
 3183

3184 **copy** Specify the name of an existing FDCC-set to be used as the source  
 3185 for the definition of this category. If this keyword is specified, no  
 3186 other keyword is specified.  
 3187 **tel\_int\_fmt** Define the appropriate representation of a telephone number for

3188 international use. The operand consists of a string, and can contain  
 3189 any combination of characters and field descriptors. In addition, the  
 3190 string can contain escape sequences defined below.  
 3191 **tel\_dom\_fmt**  
 3192 Define the appropriate representation of a telephone number for  
 3193 domestic use. The operand consists of a string, and can contain any  
 3194 combination of characters and field descriptors. In addition, the string  
 3195 can contain escape sequences defined below.  
 3196 **int\_select**  
 3197 The operand is a string with the digits used to call international  
 3198 telephone numbers.  
 3199 **int\_prefix**  
 3200 The operand is a string with the prefix used from other countries to  
 3201 call the area.

3202 The LC\_TELEPHONE category defines the interpretation of a number of escape  
 3203 sequences. The escape sequences are also available in the definitions with the following  
 3204 LC\_TELEPHONE keywords: "tel\_int\_fmt" and "tel\_dom\_fmt".  
 3205 %a area code without prefix (prefix is often <0>).  
 3206 %A area code including prefix (prefix is often <0>).  
 3207 %l local number.  
 3208 %c country code  
 3209 %C alternative carrier service code used for dialling abroad

3210 The "i18n" LC\_TELEPHONE category is:  
 3211  
 3212

```
LC_TELEPHONE
% This is the ISO/IEC TR 14652 "i18n" definition for
% the LC_TELEPHONE category.
%
tel_int_fmt    "<U002B><U0025><U0063><U0020><U002B><U0061><U0020><U002B>/
<U006C>"
```

3213 END LC\_TELEPHONE

## 3214 5. CHARMAP

3215 A character set description may exist for each coded character set supported by an  
 3216 application. This text is referred elsewhere in this Technical Report as a charmap.

3217 A conforming charmap to be used with a FDCC-set supports the portable character set  
 3218 specified in Table 1.

3219 Conforming charmaps specify certain character and character set attributes, as defined in  
 3220 5.1.

### 3221 5.1 Character Set Description Text

3222 The character set description text (charmap) describes the mapping between symbolic  
 3223 character names and actual encoding of a coded character set. It is used to bind the  
 3224 symbolic character names in a FDCC-set to an actual encoding, so an application can  
 3225 process data in this encoding.

3226 The following declarations can precede the character definitions. Each consist of the  
 3227 symbol shown in the following list, starting in column 1, including the surrounding  
 3228 brackets, followed by one or more "blank"s, followed by the value to be assigned to the  
 3229 symbol. If any of the declarations are included, they are specified in the order shown in

3243	the following list:
3244	
3245	<b>&lt;code_set_name&gt;</b>
3246	The name of the coded character set for which the character set description text is defined. The characters of the name are taken from the set of characters with visible glyphs defined in Table 1.
3247	
3248	
3249	<b>&lt;mb_cur_max&gt;</b>
3250	The maximum number of bytes in a multibyte character. This defaults to 1.
3251	
3252	<b>&lt;mb_cur_min&gt;</b>
3253	An unsigned positive integer value that defines the minimum number of bytes in a character for the encoded character set. The value is less or equal to "mb_cur_max". If not specified, the minimum number is equal to "mb_cur_max".
3254	
3255	
3256	
3257	<b>&lt;escape_char&gt;</b>
3258	The escape character used to indicate that the characters following is interpreted in a special way, as defined later in this subclause. This defaults to backslash (\). The character slash (/) is used in all the following text and examples, unless otherwise noted.
3259	
3260	
3261	
3262	
3263	<b>&lt;comment_char&gt;</b>
3264	The character that when placed in column 1 of a charmap line, is used to indicate that the line is ignored. The default character is the number sign (#). The character percent-sign (%) is used in all the following text and examples, unless otherwise noted.
3265	
3266	
3267	
3268	<b>&lt;repertoiremap&gt;</b>
3269	
3270	
3271	
3272	
3273	<b>&lt;escseq2022&gt;</b>
3274	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 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.
3275	
3276	
3277	
3278	
3279	
3280	
3281	
3282	
3283	<b>&lt;addset&gt;</b>
3284	the name of the charmap to be added to the current coded character set, and to be selected by the escape sequences defined by <escseq> of the added charmap.
3285	
3286	
3287	<b>&lt;include&gt;</b>
3288	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
3289	
3290	
3291	
3292	
3293	
3294	

3295 be included. The coded character sets are defined initially for the  
3296 encoding, and therefore do not need escape sequences for  
3297 identification. If two g0 sets are defined, the second is switched  
3298 to using the SHIFT OUT control character, while the first is  
3299 shifted to using the SHIFT IN control character.  
3300

3301 The character set mapping definitions are all the lines immediately following an identifier  
3302 line containing the string "CHARMAP" starting in column 1, and preceding a trailer line  
3303 containing the string "END CHARMAP" starting in column 1. Empty lines and lines  
3304 containing a <comment\_char> in the first column are ignored. Each noncomment line of  
3305 the character set mapping definition (i.e., between the "CHARMAP" and "END  
3306 CHARMAP" lines of the text) is in one of the following syntaxes.  
3307  
3308

3309 "%s %s %s\n", <symbolic-name>,<encoding>,<comments>  
3310

3311 "%s...%s %s %s\n", <symbolic-name>,<symbolic-name>,<encoding>,<comments>  
3312

3313 "%s....%s %s %s\n", <symbolic-name>,<symbolic-name>,<encoding>,<comments>  
3314

3315 "%s..%s %s %s\n", <symbolic-name>,<symbolic-name>,<encoding>,<comments>  
3316

3317 In the first syntax, the line of the character set mapping definition starts with the symbolic  
3318 name, immediately preceded by a <less-than> character and immediately followed by a  
3319 <greater-than> character. Symbolic names only contain characters from the set shown  
3320 with a visible glyph in Table 1.  
3321

3322 The same symbolic name may occur several times, with different values. The first value is  
3323 the one used when generating an encoding, while the other values are accepted in  
3324 decoding. Symbolic names may be included to identify values that can overlap with each  
3325 other or with the values of the symbolic names shown in Table 1. It is possible to specify  
3326 symbolic names for which no encoding exists in the encoded character set, by not  
3327 specifying a value.  
3328

3329 In the second and third syntax (symbolic decimal ellipsis), the line in the character set  
3330 mapping defines a range of one or more symbolic names. The difference between the  
3331 second and the third syntax is the number of dots in the ellipsis: the second has 3 dots, the  
3332 third has 4 dots. In these forms the symbolic names consist of zero or more nonnumeric  
3333 characters from the set shown with visible glyphs in Table 1, followed by an integer  
3334 formed by one or more decimal digits. The characters preceding the integer are identical  
3335 in the two symbolic names, and the integer formed by the digits in the second symbolic  
3336 name are identical to or greater than the integer formed by the digits in the first name.  
3337 This is interpreted as a series of symbolic names formed from the common part and each  
3338 of the integers in decimal format between the first and the second integer, inclusive, and  
3339 with a length of the symbolic names generated that is equal to the length of the first (and  
3340 also the second) symbolic name. As an example, <j0101>....<j0104> is interpreted as the  
3341 symbolic names <j0101>, <j0102>, <j0103>, and <j0104>, in that order.  
3342

3343 Note: The rationale to allow both a 3-dot and a 4-dot symbol for symbolic decimal  
3344 ellipses is that in the POSIX standard the decimal symbolic ellipses was defined by a 3-  
3345 dot symbol for charmaps, while the 3-dot symbol was an absolute ellipses for POSIX  
3346 locales, and this Technical Report specifies a 4-dot symbol for the decimal symbolic

3347     ellipses. The 3-dot symbolic decimal ellipses in charmaps is deprecated.  
3348

3349     In the fourth syntax (symbolic hexadecimal ellipsis, with two dots), the line in the  
3350     character set mapping defines a range of one or more symbolic names. In this form the  
3351     symbolic names consist of zero or more nonnumeric characters from the set shown with  
3352     visible glyphs in Table 1, followed by an integer formed by one or more hexadecimal  
3353     digits, using uppercase letters only for the range "A" to "F". The characters preceding the  
3354     hexadecimal integer are identical in the two symbolic names, and the integer formed by  
3355     the hexadecimal digits in the second symbolic name is identical to or greater than the  
3356     integer formed by the hexadecimal digits in the first name. This is interpreted as a series  
3357     of symbolic names formed from the common part and each of the integers in hexadecimal  
3358     format using uppercase letters only between the first and the second integer, inclusive, and  
3359     with a length of the symbolic names generated that is equal to the length of the first (and  
3360     also the second) symbolic name. As an example, <U010E>..<U0111> is interpreted as the  
3361     symbolic names <U010E>, <U010F>, <U0110>, and <U0111>, in that order.  
3362

3363     The encoding part is expressed as one (for single-byte values) or more concatenated  
3364     decimal, octal or hexadecimal constants (hexadecimal constants is recommended). Decimal  
3365     constants are represented by two or three decimal digits, preceded by the escape character  
3366     and the lowercase letter "d"; for example /d05, /d97, or /d143. Hexadecimal constants are  
3367     represented by two hexadecimal digits, preceded by the escape character and the lowercase  
3368     letter "x"; for example /x05, /x61, or /x8f. Octal constants are represented by two or three  
3369     octal digits, preceded by the escape character; for example /05, /141, or /217. In a  
3370     charmap, each constant should represent an 8 bit byte for portability reasons. Applications  
3371     supporting other byte sizes may allow constants to represent values larger than those that  
3372     can be represented in 8 bit bytes, and to allow additional digits in constants. When  
3373     constants are concatenated for multibyte character values, they may be of different types,  
3374     and interpreted in byte order from the first to the last with the least significant byte of the  
3375     multibyte character specified by the last byte. The manner in which these constants are  
3376     represented in the character stored in the system is application defined. Omitting bytes  
3377     from a multibyte character produces undefined results.  
3378

3379     In lines defining ranges of symbolic names, the encoded value is the value for the first  
3380     symbolic name in the range (the symbolic name preceding the ellipsis). Subsequent  
3381     symbolic names defined by the range have encoding values in increasing order. For  
3382     example the line  
3383

3384        <j0101>....<j0104>           /d129/d254  
3385

3386     is interpreted as  
3387

3388        <j0101>    /d129/d254  
3389        <j0102>    /d129/d255  
3390        <j0103>    /d130/d000  
3391        <j0104>    /d130/d001  
3392

3393     The comments parameter is optional.  
3394

3395     Example of using ISO 2022 techniques:  
3396

3397     The following example defines two coded character sets, a 7-bit and a 14-bit. They are then merged into one  
3398     encoding. It is an example on how encodings used in Eastern Asia could be specified.  
3399

3400 The 7-bit charmap  
 3401  
 3402 <escape\_char> /  
 3403 <comment\_char> %  
 3404 % The 7bit charmap defines both control and graphic characters  
 3405 <code\_set\_name> "eastern7bit"  
 3406 <escseq2022> "c0"; "c0", "/x21/x40"  
 3407 <escseq2022> "g0"; "g0", "/x28/x48"  
 3408 <escseq2022> "g1"; "g0", "/x29/x48"  
 3409 <escseq2022> "g2"; "g0", "/x2A/x48"  
 3410 <escseq2022> "g3"; "g0", "/x2B/x48"  
 3411  
 3412 CHARMAP  
 3413 <tab> /x08  
 3414 <newline> /x0D  
 3415 <a> /x61  
 3416 % more character encodings to be defined here  
 3417 END CHARMAP  
 3418  
 3419

## The 14-bit charmap

3420  
 3421 <escape\_char> /  
 3422 <comment\_char> %  
 3423 <code\_set\_name> "eastern14bit"  
 3424 <mb\_cur\_max> 2  
 3425 <esqseq> "g0"; "g0", "/x24/x40"  
 3426 <esqseq> "g1"; "g0", "/x24/x29/x40"  
 3427 <esqseq> "g2"; "g0", "/x24/x2A/x40"  
 3428 <esqseq> "g3"; "g0", "/x24/x2B/x40"  
 3429  
 3430 CHARMAP  
 3431 <U0165> /d036/d055 % the character codes are only examples  
 3432 <U0244> /d036/d056  
 3433 % more character encodings to be defined here  
 3434 END CHARMAP  
 3435  
 3436

## The merged encoding

3437  
 3438 <escape\_char> /  
 3439 <comment\_char> %  
 3440 <code\_set\_name> "shift-eastern"  
 3441 <mb\_cur\_max> 2  
 3442 <mb\_cur\_min> 1  
 3443 <include> "c0"; "c0"; "eastern7bit"  
 3444 <include> "g0"; "g0"; "eastern7bit"  
 3445 <include> "g1"; "g0"; "eastern14bit"  
 3446  
 3447 % This defines the g0 values of "eastern14bit" (without the 8th  
 3448 % bit set) to be the g1 in this encoding (with the 8th bit set).  
 3449 %  
 3450 % So the bytes without the 8th bit set is from the "shift7bit"  
 3451 % coded character set, while bytes with the 8th bit set are from  
 3452 % the 14-bit set.  
 3453

## Another merged encoding using the same charmaps:

3454  
 3455 <escape\_char> /  
 3456 <comment\_char> %  
 3457 <code\_set\_name> "EUC-eastern"  
 3458 <mb\_cur\_max> 2  
 3459 <mb\_cur\_min> 1  
 3460 <include> "c0"; "c0"; "eastern7bit"  
 3461 <include> "g0"; "g0"; "eastern7bit"  
 3462 <include> "g0"; "g0"; "eastern14bit"  
 3463  
 3464 % As there are two "g0" sets defined, the first referenced is the  
 3465 % initial g0 set, while the second can be shifted to via the SHIFT OUT  
 3466 % control character. The first can then be shifted to by the SHIFT IN  
 3467 % control character.  
 3468  
 3469

3470   **WIDTH section**

3471  
 3472   After the "END CHARMAP" statement the following declarations may follow. Each  
 3473   consists of the keyword shown in the following list, starting in column 1, followed by the  
 3474   value(s) to be associated to the keyword, as defined below.

3475  
 3476   **WIDTH** An unassigned positive integer value defining the column width for the  
 3477   characters in the coded character set. Coded character values are defined using symbolic  
 3478   character names followed by a column width value. Defining a character with more than  
 3479   one WIDTH produces undefined results. The END WIDTH keyword is used to terminate  
 3480   the WIDTH definitions.

3481  
 3482   **WIDTH\_DEFAULT** An unsigned positive integer value defining the column width for  
 3483   any character not listed by one of the WIDTH keywords. If no WIDTH\_DEFAULT  
 3484   keyword is included in the charmap, the default character width is 1.

3485  
 3486   Example:

3487  
 3488   After the "END CHARMAP" statement, a syntax for width definition would be:

3489  
 3490   **WIDTH**  
 3491   <A> 1  
 3492   <B> 1  
 3493   <j0101>...<j0195> 2  
 3494   <U3200>..<UFAFF> 2  
 3495   **END WIDTH**  
 3496   **WIDTH\_DEFAULT** 1  
 3497

3498   In this example, the code point values represented by <A> and <B> are assigned a width  
 3499   of 1. The code point values <j0101>...<j0195> (decimal ellipses) and <U3200>..<UFAFF>  
 3500   are assigned a width of 2. The last line defines the DEFAULT\_WIDTH to 1.

3501  
 3502   **6 REPERTOIREMAP**

3503  
 3504   FDCC-set and Charmap sources may be specified in a coded character set independent  
 3505   way, using symbolic character names. The relation between the symbolic character names  
 3506   and characters may be specified via a Repertoiremap, which defines the repertoire of  
 3507   characters defined for a FDCC-set, and the symbolic character names and corresponding  
 3508   abstract character (by a reference to ISO/IEC 10646).

3509  
 3510   The repertoire mapping is defined by specifying the symbolic character name and the  
 3511   ISO/IEC 10646 code position in hexadecimal form (with a preceding 'U') and optionally  
 3512   the long ISO/IEC 10646 character name in the following syntax:

3513   "%s %s %s\n",<symbolic-name>,<10646-short-identifier>,<comments>

3514  
 3515   The symbolic character name and the ISO/IEC 10646 short identifier are each surrounded  
 3516   by angle brackets <>, and the fields are separated by one or more spaces or tabs on a line.  
 3517   If a right angle bracket or an escape character is used within a symbolic name, it is  
 3518   preceded by the escape character. Characters not in ISO/IEC 10646 may be referenced by  
 3519   the symbolic character names <P0000000>..<PF8FFFFFFF>.

3522 The escape character can be redefined from the default reverse solidus () with the first  
 3523 line of the Repertoiremap containing the string "escape\_char" followed by one or more  
 3524 spaces or tabs and then the escape character.

3525

3526 Several symbolic character names can refer to the same abstract character, and are then  
 3527 used as synonyms in FDCC-sets and charmaps. The set of <U0000>..<UFFFF> and  
 3528 <U00000000>..<U7FFFFFF> symbolic names (no lowercase letters) are predefined and  
 3529 refers to the corresponding code points of ISO/IEC 10646 with the same short identifier.

3530

3531 The "i18nrep" repertoiremap is defined to accommodate prior art, such as defined in  
 3532 Annex G of the ISO/IEC 9945-2:1993 standard, and used by ISO and IEC member bodies  
 3533 in their national POSIX locale specifications, and as used in POSIX locales distributed by  
 3534 the ISO/IEC POSIX working group and The Open Group. Many POSIX charmaps  
 3535 registered with ISO/IEC 15897 use these symbolic names. It also reflects use on the  
 3536 Internet, and many of the Internet registered charsets are specified using these symbolic  
 3537 names. The "i18nrep" repertoiremap thus facilitates reuse of both POSIX locale data and  
 3538 POSIX charmaps with data from this Technical Report. The sequence <a8>..<z8> are used  
 3539 as hooks for tailoring to denote the last accented Latin letter of each of the ISO/IEC 646  
 3540 letters <a>..<z>, so that tailorings that need to have specifications after the last letter of  
 3541 such a family, for example to introduce a new letter of an alphabet, can do so with a  
 3542 reference that is stable over different versions of the "i18n" FDCC-set. The contents of the  
 3543 "i18nrep" repertoiremap is as follows:

3544

```

3545 escape_char /
3546 <NUL>           <U0000>  NULL (NUL)
3547 <SOH>           <U0001>  START OF HEADING (SOH)
3548 <STX>           <U0002>  START OF TEXT (STX)
3549 <ETX>           <U0003>  END OF TEXT (ETX)
3550 <EOT>           <U0004>  END OF TRANSMISSION (EOT)
3551 <ENQ>           <U0005>  ENQUIRY (ENQ)
3552 <ACK>           <U0006>  ACKNOWLEDGE (ACK)
3553 <alert>          <U0007>  BELL (BEL)
3554 <BEL>           <U0007>  BELL (BEL)
3555 <backspace>     <U0008>  BACKSPACE (BS)
3556 <tab>            <U0009>  CHARACTER TABULATION (HT)
3557 <newline>        <U000A>  LINE FEED (LF)
3558 <vertical-tab>   <U000B>  LINE TABULATION (VT)
3559 <form-feed>      <U000C>  FORM FEED (FF)
3560 <carriage-return> <U000D>  CARRIAGE RETURN (CR)
3561 <DLE>            <U0010>  DATALINK ESCAPE (DLE)
3562 <DC1>            <U0011>  DEVICE CONTROL ONE (DC1)
3563 <DC2>            <U0012>  DEVICE CONTROL TWO (DC2)
3564 <DC3>            <U0013>  DEVICE CONTROL THREE (DC3)
3565 <DC4>            <U0014>  DEVICE CONTROL FOUR (DC4)
3566 <NAK>            <U0015>  NEGATIVE ACKNOWLEDGE (NAK)
3567 <SYN>            <U0016>  SYNCHRONOUS IDLE (SYN)
3568 <ETB>            <U0017>  END OF TRANSMISSION BLOCK (ETB)
3569 <CAN>            <U0018>  CANCEL (CAN)
3570 <SUB>            <U001A>  SUBSTITUTE (SUB)
3571 <ESC>            <U001B>  ESCAPE (ESC)
3572 <IS4>             <U001C>  FILE SEPARATOR (IS4)
3573 <IS3>             <U001D>  GROUP SEPARATOR (IS3)
3574 <intro>          <U001D>  GROUP SEPARATOR (IS3)
3575 <IS2>             <U001E>  RECORD SEPARATOR (IS2)
3576 <IS1>             <U001F>  UNIT SEPARATOR (IS1)
3577 <DEL>             <U007F>  DELETE (DEL)
3578 <space>           <U0020>  SPACE
3579 <exclamation-mark> <U0021>  EXCLAMATION MARK
3580 <quotation-mark>  <U0022>  QUOTATION MARK
3581 <number-sign>     <U0023>  NUMBER SIGN
3582 <dollar-sign>    <U0024>  DOLLAR SIGN
3583 <percent-sign>   <U0025>  PERCENT SIGN
3584 <ampersand>      <U0026>  AMPERSAND
3585 <apostrophe>     <U0027>  APOSTROPHE
3586 <left-parenthesis> <U0028>  LEFT PARENTHESIS
3587 <right-parenthesis> <U0029>  RIGHT PARENTHESIS
3588 <asterisk>        <U002A>  ASTERISK
3589 <plus-sign>       <U002B>  PLUS SIGN
3590 <comma>           <U002C>  COMMA
3591 <hyphen>          <U002D>  HYPHEN-MINUS
3592 <hyphen-minus>   <U002D>  HYPHEN-MINUS
3593 <period>          <U002E>  FULL STOP
3594 <full-stop>       <U002E>  FULL STOP
  
```

3595	<slash>	<U002F> SOLIDUS
3596	<solidus>	<U002F> SOLIDUS
3597	<zero>	<U0030> DIGIT ZERO
3598	<one>	<U0031> DIGIT ONE
3599	<two>	<U0032> DIGIT TWO
3600	<three>	<U0033> DIGIT THREE
3601	<four>	<U0034> DIGIT FOUR
3602	<five>	<U0035> DIGIT FIVE
3603	<six>	<U0036> DIGIT SIX
3604	<seven>	<U0037> DIGIT SEVEN
3605	<eight>	<U0038> DIGIT EIGHT
3606	<nine>	<U0039> DIGIT NINE
3607	<colon>	<U003A> COLON
3608	&ltsemicolon>	<U003B> SEMICOLON
3609	<less-than-sign>	<U003C> LESS-THAN SIGN
3610	<equals-sign>	<U003D> EQUALS SIGN
3611	<greater-than-sign>	<U003E> GREATER-THAN SIGN
3612	<question-mark>	<U003F> QUESTION MARK
3613	<commercial-at>	<U0040> COMMERCIAL AT
3614	<left-square-bracket>	<U005B> LEFT SQUARE BRACKET
3615	<backslash>	<U005C> REVERSE SOLIDUS
3616	<reverse-solidus>	<U005C> REVERSE SOLIDUS
3617	<right-square-bracket>	<U005D> RIGHT SQUARE BRACKET
3618	<circumflex>	<U005E> CIRCUMFLEX ACCENT
3619	<circumflex-accent>	<U005E> CIRCUMFLEX ACCENT
3620	<underscore>	<U005F> LOW LINE
3621	<low-line>	<U005F> LOW LINE
3622	<grave-accent>	<U0060> GRAVE ACCENT
3623	<left-brace>	<U007B> LEFT CURLY BRACKET
3624	<left-curly-bracket>	<U007B> LEFT CURLY BRACKET
3625	<vertical-line>	<U007C> VERTICAL LINE
3626	<right-brace>	<U007D> RIGHT CURLY BRACKET
3627	<right-curly-bracket>	<U007D> RIGHT CURLY BRACKET
3628	<tilde>	<U007E> TILDE
3629	<a8>	<U0252> Weight indicating the position of the last a
3630	<b8>	<U0182> Weight indicating the position of the last b
3631	<c8>	<U0255> Weight indicating the position of the last c
3632	<d8>	<U018D> Weight indicating the position of the last d
3633	<e8>	<U0264> Weight indicating the position of the last e
3634	<f8>	<U0191> Weight indicating the position of the last f
3635	<g8>	<U01A2> Weight indicating the position of the last g
3636	<h8>	<U02BD> Weight indicating the position of the last h
3637	<i8>	<U0196> Weight indicating the position of the last i
3638	<j8>	<U0284> Weight indicating the position of the last j
3639	<k8>	<U029E> Weight indicating the position of the last k
3640	<l8>	<U028E> Weight indicating the position of the last l
3641	<m8>	<U0271> Weight indicating the position of the last m
3642	<n8>	<U014A> Weight indicating the position of the last n
3643	<o8>	<U0277> Weight indicating the position of the last o
3644	<p8>	<U0278> Weight indicating the position of the last p
3645	<q8>	<U0138> Weight indicating the position of the last q
3646	<r8>	<U02B6> Weight indicating the position of the last r
3647	<s8>	<U0286> Weight indicating the position of the last s
3648	<t8>	<U0287> Weight indicating the position of the last t
3649	<u8>	<U01B1> Weight indicating the position of the last u
3650	<v8>	<U028C> Weight indicating the position of the last v
3651	<w8>	<U028D> Weight indicating the position of the last w
3652	<x8>	<U216B> Weight indicating the position of the last x
3653	<y8>	<U01B3> Weight indicating the position of the last y
3654	<z8>	<U0293> Weight indicating the position of the last z
3655	<NU>	<U0000> NULL (NUL)
3656	<SH>	<U0001> START OF HEADING (SOH)
3657	<SX>	<U0002> START OF TEXT (STX)
3658	<EX>	<U0003> END OF TEXT (ETX)
3659	<ET>	<U0004> END OF TRANSMISSION (EOT)
3660	<EQ>	<U0005> ENQUIRY (ENQ)
3661	<AK>	<U0006> ACKNOWLEDGE (ACK)
3662	<BL>	<U0007> BELL (BEL)
3663	<BS>	<U0008> BACKSPACE (BS)
3664	<HT>	<U0009> CHARACTER TABULATION (HT)
3665	<LF>	<U000A> LINE FEED (LF)
3666	<VT>	<U000B> LINE TABULATION (VT)
3667	<FF>	<U000C> FORM FEED (FF)
3668	<CR>	<U000D> CARRIAGE RETURN (CR)
3669	<SO>	<U000E> SHIFT OUT (SO)
3670	<SI>	<U000F> SHIFT IN (SI)
3671	<DL>	<U0010> DATALINK ESCAPE (DLE)
3672	<D1>	<U0011> DEVICE CONTROL ONE (DC1)
3673	<D2>	<U0012> DEVICE CONTROL TWO (DC2)
3674	<D3>	<U0013> DEVICE CONTROL THREE (DC3)
3675	<D4>	<U0014> DEVICE CONTROL FOUR (DC4)
3676	<NK>	<U0015> NEGATIVE ACKNOWLEDGE (NAK)
3677	<SY>	<U0016> SYNCHRONOUS IDLE (SYN)
3678	<EB>	<U0017> END OF TRANSMISSION BLOCK (ETB)
3679	<CN>	<U0018> CANCEL (CAN)
3680	<EM>	<U0019> END OF MEDIUM (EM)

3683	<SB>	<U001A>	SUBSTITUTE (SUB)
3684	<EC>	<U001B>	ESCAPE (ESC)
3685	<FS>	<U001C>	FILE SEPARATOR (IS4)
3686	<GS>	<U001D>	GROUP SEPARATOR (IS3)
3687	<RS>	<U001E>	RECORD SEPARATOR (IS2)
3688	<US>	<U001F>	UNIT SEPARATOR (IS1)
3689	<DT>	<U007F>	DELETE (DEL)
3690	<PA>	<U0080>	PADDING CHARACTER (PAD)
3691	<HO>	<U0081>	HIGH OCTET PRESET (HOP)
3692	<BH>	<U0082>	BREAK PERMITTED HERE (BPH)
3693	<NH>	<U0083>	NO BREAK HERE (NBH)
3694	<IN>	<U0084>	INDEX (IND)
3695	<NL>	<U0085>	NEXT LINE (NEL)
3696	<SA>	<U0086>	START OF SELECTED AREA (SSA)
3697	<ES>	<U0087>	END OF SELECTED AREA (ESA)
3698	<HS>	<U0088>	CHARACTER TABULATION SET (HTS)
3699	<HJ>	<U0089>	CHARACTER TABULATION WITH JUSTIFICATION (HTJ)
3700	<VS>	<U008A>	LINE TABULATION SET (VTS)
3701	<PD>	<U008B>	PARTIAL LINE FORWARD (PLD)
3702	<PU>	<U008C>	PARTIAL LINE BACKWARD (PLU)
3703	<RI>	<U008D>	REVERSE LINE FEED (RI)
3704	<S2>	<U008E>	SINGLE-SHIFT TWO (SS2)
3705	<S3>	<U008F>	SINGLE-SHIFT THREE (SS3)
3706	<DC>	<U0090>	DEVICE CONTROL STRING (DCS)
3707	<P1>	<U0091>	PRIVATE USE ONE (PU1)
3708	<P2>	<U0092>	PRIVATE USE TWO (PU2)
3709	<TS>	<U0093>	SET TRANSMIT STATE (STS)
3710	<CC>	<U0094>	CANCEL CHARACTER (CCH)
3711	<MW>	<U0095>	MESSAGE WAITING (MW)
3712	<SG>	<U0096>	START OF GUARDED AREA (SPA)
3713	<EG>	<U0097>	END OF GUARDED AREA (EPA)
3714	<SS>	<U0098>	START OF STRING (SOS)
3715	<GC>	<U0099>	SINGLE GRAPHIC CHARACTER INTRODUCER (SGCI)
3716	<SC>	<U009A>	SINGLE CHARACTER INTRODUCER (SCI)
3717	<CI>	<U009B>	CONTROL SEQUENCE INTRODUCER (CSI)
3718	<ST>	<U009C>	STRING TERMINATOR (ST)
3719	<OC>	<U009D>	OPERATING SYSTEM COMMAND (OSC)
3720	<PM>	<U009E>	PRIVACY MESSAGE (PM)
3721	<AC>	<U009F>	APPLICATION PROGRAM COMMAND (APC)
3722	<SP>	<U0020>	SPACE
3723	<!>	<U0021>	EXCLAMATION MARK
3724	<">	<U0022>	QUOTATION MARK
3725	<Nb>	<U0023>	NUMBER SIGN
3726	<DO>	<U0024>	DOLLAR SIGN
3727	<%>	<U0025>	PERCENT SIGN
3728	<&>	<U0026>	AMPERSAND
3729	<'>	<U0027>	APOSTROPHE
3730	<(>	<U0028>	LEFT PARENTHESIS
3731	<)>	<U0029>	RIGHT PARENTHESIS
3732	<*>	<U002A>	ASTERISK
3733	<+>	<U002B>	PLUS SIGN
3734	<,>	<U002C>	COMMA
3735	<->	<U002D>	HYPHEN-MINUS
3736	<. >	<U002E>	FULL STOP
3737	</ />	<U002F>	SOLIDUS
3738	<0>	<U0030>	DIGIT ZERO
3739	<1>	<U0031>	DIGIT ONE
3740	<2>	<U0032>	DIGIT TWO
3741	<3>	<U0033>	DIGIT THREE
3742	<4>	<U0034>	DIGIT FOUR
3743	<5>	<U0035>	DIGIT FIVE
3744	<6>	<U0036>	DIGIT SIX
3745	<7>	<U0037>	DIGIT SEVEN
3746	<8>	<U0038>	DIGIT EIGHT
3747	<9>	<U0039>	DIGIT NINE
3748	<:>	<U003A>	COLON
3749	<;>	<U003B>	SEMICOLON
3750	<<>	<U003C>	LESS-THAN SIGN
3751	<=>	<U003D>	EQUALS SIGN
3752	</>>	<U003E>	GREATER-THAN SIGN
3753	<?>	<U003F>	QUESTION MARK
3754	<AT>	<U0040>	COMMERCIAL AT
3755	<A>	<U0041>	LATIN CAPITAL LETTER A
3756	<B>	<U0042>	LATIN CAPITAL LETTER B
3757	<C>	<U0043>	LATIN CAPITAL LETTER C
3758	<D>	<U0044>	LATIN CAPITAL LETTER D
3759	<E>	<U0045>	LATIN CAPITAL LETTER E
3760	<F>	<U0046>	LATIN CAPITAL LETTER F
3761	<G>	<U0047>	LATIN CAPITAL LETTER G
3762	<H>	<U0048>	LATIN CAPITAL LETTER H
3763	<I>	<U0049>	LATIN CAPITAL LETTER I
3764	<J>	<U004A>	LATIN CAPITAL LETTER J
3765	<K>	<U004B>	LATIN CAPITAL LETTER K
3766	<L>	<U004C>	LATIN CAPITAL LETTER L
3767	<M>	<U004D>	LATIN CAPITAL LETTER M
3768	<N>	<U004E>	LATIN CAPITAL LETTER N
3769	<O>	<U004F>	LATIN CAPITAL LETTER O
3770	<P>	<U0050>	LATIN CAPITAL LETTER P

3771	<Q>	<U0051>	LATIN CAPITAL LETTER Q
3772	<R>	<U0052>	LATIN CAPITAL LETTER R
3773	<S>	<U0053>	LATIN CAPITAL LETTER S
3774	<T>	<U0054>	LATIN CAPITAL LETTER T
3775	<U>	<U0055>	LATIN CAPITAL LETTER U
3776	<V>	<U0056>	LATIN CAPITAL LETTER V
3777	<W>	<U0057>	LATIN CAPITAL LETTER W
3778	<X>	<U0058>	LATIN CAPITAL LETTER X
3779	<Y>	<U0059>	LATIN CAPITAL LETTER Y
3780	<Z>	<U005A>	LATIN CAPITAL LETTER Z
3781	<<(>	<U005B>	LEFT SQUARE BRACKET
3782	<////>	<U005C>	REVERSE SOLIDUS
3783	<)/>>	<U005D>	RIGHT SQUARE BRACKET
3784	<' />>	<U005E>	CIRCUMFLEX ACCENT
3785	<_>	<U005F>	LOW LINE
3786	<' !>	<U0060>	GRAVE ACCENT
3787	<a>	<U0061>	LATIN SMALL LETTER A
3788	<b>	<U0062>	LATIN SMALL LETTER B
3789	<c>	<U0063>	LATIN SMALL LETTER C
3790	<d>	<U0064>	LATIN SMALL LETTER D
3791	<e>	<U0065>	LATIN SMALL LETTER E
3792	<f>	<U0066>	LATIN SMALL LETTER F
3793	<g>	<U0067>	LATIN SMALL LETTER G
3794	<h>	<U0068>	LATIN SMALL LETTER H
3795	<i>	<U0069>	LATIN SMALL LETTER I
3796	<j>	<U006A>	LATIN SMALL LETTER J
3797	<k>	<U006B>	LATIN SMALL LETTER K
3798	<l>	<U006C>	LATIN SMALL LETTER L
3799	<m>	<U006D>	LATIN SMALL LETTER M
3800	<n>	<U006E>	LATIN SMALL LETTER N
3801	<o>	<U006F>	LATIN SMALL LETTER O
3802	<p>	<U0070>	LATIN SMALL LETTER P
3803	<q>	<U0071>	LATIN SMALL LETTER Q
3804	<r>	<U0072>	LATIN SMALL LETTER R
3805	<s>	<U0073>	LATIN SMALL LETTER S
3806	<t>	<U0074>	LATIN SMALL LETTER T
3807	<u>	<U0075>	LATIN SMALL LETTER U
3808	<v>	<U0076>	LATIN SMALL LETTER V
3809	<w>	<U0077>	LATIN SMALL LETTER W
3810	<x>	<U0078>	LATIN SMALL LETTER X
3811	<y>	<U0079>	LATIN SMALL LETTER Y
3812	<z>	<U007A>	LATIN SMALL LETTER Z
3813	<(>	<U007B>	LEFT CURLY BRACKET
3814	<!>	<U007C>	VERTICAL LINE
3815	<!>	<U007D>	RIGHT CURLY BRACKET
3816	<' ?>	<U007E>	TIFFE
3817	<NS>	<U00A0>	NO-BREAK SPACE
3818	<!I>	<U00A1>	INVERTED EXCLAMATION MARK
3819	<Ct>	<U00A2>	CENT SIGN
3820	<Pd>	<U00A3>	POUND SIGN
3821	<Cu>	<U00A4>	CURRENCY SIGN
3822	<Ye>	<U00A5>	YEN SIGN
3823	<BB>	<U00A6>	BROKEN BAR
3824	<SE>	<U00A7>	SECTION SIGN
3825	<':>	<U00A8>	DIAERESIS
3826	<Co>	<U00A9>	COPYRIGHT SIGN
3827	<-a>	<U00AA>	FEMININE ORDINAL INDICATOR
3828	<<<>	<U00AB>	LEFT-POINTING DOUBLE ANGLE QUOTATION MARK
3829	<NO>	<U00AC>	NOT SIGN
3830	<-->	<U00AD>	SOFT HYPHEN
3831	<Rg>	<U00AE>	REGISTERED SIGN
3832	<' m>	<U00AF>	MACRON
3833	<DG>	<U00B0>	DEGREE SIGN
3834	<+->	<U00B1>	PLUS-MINUS SIGN
3835	<2S>	<U00B2>	SUPERSCRIPT TWO
3836	<3S>	<U00B3>	SUPERSCRIPT THREE
3837	<' '>	<U00B4>	ACUTE ACCENT
3838	<My>	<U00B5>	MICRO SIGN
3839	<PI>	<U00B6>	PILCROW SIGN
3840	<.M>	<U00B7>	MIDDLE DOT
3841	<',>	<U00B8>	CEDILLA
3842	<1S>	<U00B9>	SUPERSCRIPT ONE
3843	<-o>	<U00BA>	MASCULINE ORDINAL INDICATOR
3844	</>/>>	<U00BB>	RIGHT-POINTING DOUBLE ANGLE QUOTATION MARK
3845	<14>	<U00BC>	VULGAR FRACTION ONE QUARTER
3846	<12>	<U00BD>	VULGAR FRACTION ONE HALF
3847	<34>	<U00BE>	VULGAR FRACTION THREE QUARTERS
3848	<?I>	<U00BF>	INVERTED QUESTION MARK
3849	<A!>	<U00C0>	LATIN CAPITAL LETTER A WITH GRAVE
3850	<A'>	<U00C1>	LATIN CAPITAL LETTER A WITH ACUTE
3851	<A/>>	<U00C2>	LATIN CAPITAL LETTER A WITH CIRCUMFLEX
3852	<A?>	<U00C3>	LATIN CAPITAL LETTER A WITH TILDE
3853	<A:>	<U00C4>	LATIN CAPITAL LETTER A WITH DIAERESIS
3854	<AA>	<U00C5>	LATIN CAPITAL LETTER A WITH RING ABOVE
3855	<AE>	<U00C6>	LATIN CAPITAL LETTER AE (ash)
3856	<C,>	<U00C7>	LATIN CAPITAL LETTER C WITH CEDILLA
3857	<E!>	<U00C8>	LATIN CAPITAL LETTER E WITH GRAVE
3858	<E'>	<U00C9>	LATIN CAPITAL LETTER E WITH ACUTE

3850	<E/>	<U00CA>	LATIN CAPITAL LETTER E WITH CIRCUMFLEX
3860	<E:>	<U00CB>	LATIN CAPITAL LETTER E WITH DIAERESIS
3861	<I/>	<U00CC>	LATIN CAPITAL LETTER I WITH GRAVE
3862	<I'>	<U00CD>	LATIN CAPITAL LETTER I WITH ACUTE
3863	<I//>	<U00CE>	LATIN CAPITAL LETTER I WITH CIRCUMFLEX
3864	<I:>	<U00CF>	LATIN CAPITAL LETTER I WITH DIAERESIS
3865	<D->	<U00D0>	LATIN CAPITAL LETTER ETH (Icelandic)
3866	<N?>	<U00D1>	LATIN CAPITAL LETTER N WITH TILDE
3867	<O!>	<U00D2>	LATIN CAPITAL LETTER O WITH GRAVE
3868	<O'>	<U00D3>	LATIN CAPITAL LETTER O WITH ACUTE
3869	<O//>	<U00D4>	LATIN CAPITAL LETTER O WITH CIRCUMFLEX
3870	<O?>	<U00D5>	LATIN CAPITAL LETTER O WITH TILDE
3871	<O:>	<U00D6>	LATIN CAPITAL LETTER O WITH DIAERESIS
3872	<X>	<U00D7>	MULTIPLICATION SIGN
3873	<O//>	<U00D8>	LATIN CAPITAL LETTER O WITH STROKE
3874	<U!>	<U00D9>	LATIN CAPITAL LETTER U WITH GRAVE
3875	<U'>	<U00DA>	LATIN CAPITAL LETTER U WITH ACUTE
3876	<U//>	<U00DB>	LATIN CAPITAL LETTER U WITH CIRCUMFLEX
3877	<U:>	<U00DC>	LATIN CAPITAL LETTER U WITH DIAERESIS
3878	<Y'>	<U00DD>	LATIN CAPITAL LETTER Y WITH ACUTE
3879	<TH>	<U00DE>	LATIN CAPITAL LETTER THORN (Icelandic)
3880	<ss>	<U00DF>	LATIN SMALL LETTER SHARP S (German)
3881	<a!>	<U00E0>	LATIN SMALL LETTER A WITH GRAVE
3882	<a'>	<U00E1>	LATIN SMALL LETTER A WITH ACUTE
3883	<a//>	<U00E2>	LATIN SMALL LETTER A WITH CIRCUMFLEX
3884	<a?>	<U00E3>	LATIN SMALL LETTER A WITH TILDE
3885	<a:>	<U00E4>	LATIN SMALL LETTER A WITH DIAERESIS
3886	<aa>	<U00E5>	LATIN SMALL LETTER A WITH RING ABOVE
3887	<ae>	<U00E6>	LATIN SMALL LETTER AE (ash)
3888	<c,>	<U00E7>	LATIN SMALL LETTER C WITH CEDILLA
3889	<e!>	<U00E8>	LATIN SMALL LETTER E WITH GRAVE
3890	<e'>	<U00E9>	LATIN SMALL LETTER E WITH ACUTE
3891	<e//>	<U00EA>	LATIN SMALL LETTER E WITH CIRCUMFLEX
3892	<e:>	<U00EB>	LATIN SMALL LETTER E WITH DIAERESIS
3893	<i!>	<U00EC>	LATIN SMALL LETTER I WITH GRAVE
3894	<i'>	<U00ED>	LATIN SMALL LETTER I WITH ACUTE
3895	<i//>	<U00EE>	LATIN SMALL LETTER I WITH CIRCUMFLEX
3896	<i:>	<U00EF>	LATIN SMALL LETTER I WITH DIAERESIS
3897	<d->	<U00F0>	LATIN SMALL LETTER ETH (Icelandic)
3898	<n?>	<U00F1>	LATIN SMALL LETTER N WITH TILDE
3899	<o!>	<U00F2>	LATIN SMALL LETTER O WITH GRAVE
3900	<o'>	<U00F3>	LATIN SMALL LETTER O WITH ACUTE
3901	<o//>	<U00F4>	LATIN SMALL LETTER O WITH CIRCUMFLEX
3902	<o?>	<U00F5>	LATIN SMALL LETTER O WITH TILDE
3903	<o:>	<U00F6>	LATIN SMALL LETTER O WITH DIAERESIS
3904	<-:>	<U00F7>	DIVISION SIGN
3905	<o//>	<U00F8>	LATIN SMALL LETTER O WITH STROKE
3906	<u!>	<U00F9>	LATIN SMALL LETTER U WITH GRAVE
3907	<u'>	<U00FA>	LATIN SMALL LETTER U WITH ACUTE
3908	<u//>	<U00FB>	LATIN SMALL LETTER U WITH CIRCUMFLEX
3909	<u:>	<U00FC>	LATIN SMALL LETTER U WITH DIAERESIS
3910	<y'>	<U00FD>	LATIN SMALL LETTER Y WITH ACUTE
3911	<th>	<U00FE>	LATIN SMALL LETTER THORN (Icelandic)
3912	<y:>	<U00FF>	LATIN SMALL LETTER Y WITH DIAERESIS
3913	<A->	<U0100>	LATIN CAPITAL LETTER A WITH MACRON
3914	<a->	<U0101>	LATIN SMALL LETTER A WITH MACRON
3915	<A(>	<U0102>	LATIN CAPITAL LETTER A WITH BREVE
3916	<a(>	<U0103>	LATIN SMALL LETTER A WITH BREVE
3917	<A;>	<U0104>	LATIN CAPITAL LETTER A WITH OGONEK
3918	<a;>	<U0105>	LATIN SMALL LETTER A WITH OGONEK
3919	<C'>	<U0106>	LATIN CAPITAL LETTER C WITH ACUTE
3920	<c'>	<U0107>	LATIN SMALL LETTER C WITH ACUTE
3921	<C//>	<U0108>	LATIN CAPITAL LETTER C WITH CIRCUMFLEX
3922	<c//>	<U0109>	LATIN SMALL LETTER C WITH CIRCUMFLEX
3923	<C.>	<U010A>	LATIN CAPITAL LETTER C WITH DOT ABOVE
3924	<c.>	<U010B>	LATIN SMALL LETTER C WITH DOT ABOVE
3925	<C<>	<U010C>	LATIN CAPITAL LETTER C WITH CARON
3926	<c<>	<U010D>	LATIN SMALL LETTER C WITH CARON
3927	<D->	<U010E>	LATIN CAPITAL LETTER D WITH CARON
3928	<d->	<U010F>	LATIN SMALL LETTER D WITH CARON
3929	<D//>	<U0110>	LATIN CAPITAL LETTER D WITH STROKE
3930	<d//>	<U0111>	LATIN SMALL LETTER D WITH STROKE
3931	<E->	<U0112>	LATIN CAPITAL LETTER E WITH MACRON
3932	<e->	<U0113>	LATIN SMALL LETTER E WITH MACRON
3933	<E(>	<U0114>	LATIN CAPITAL LETTER E WITH BREVE
3934	<e(>	<U0115>	LATIN SMALL LETTER E WITH BREVE
3935	<E,>	<U0116>	LATIN CAPITAL LETTER E WITH DOT ABOVE
3936	<e,>	<U0117>	LATIN SMALL LETTER E WITH DOT ABOVE
3937	<E;>	<U0118>	LATIN CAPITAL LETTER E WITH OGONEK
3938	<e;>	<U0119>	LATIN SMALL LETTER E WITH OGONEK
3939	<E<>	<U011A>	LATIN CAPITAL LETTER E WITH CARON
3940	<e<>	<U011B>	LATIN SMALL LETTER E WITH CARON
3941	<G//>	<U011C>	LATIN CAPITAL LETTER G WITH CIRCUMFLEX
3942	<g//>	<U011D>	LATIN SMALL LETTER G WITH CIRCUMFLEX
3943	<G(>	<U011E>	LATIN CAPITAL LETTER G WITH BREVE
3944	<g(>	<U011F>	LATIN SMALL LETTER G WITH BREVE
3945	<G,>	<U0120>	LATIN CAPITAL LETTER G WITH DOT ABOVE
3946	<g,>	<U0121>	LATIN SMALL LETTER G WITH DOT ABOVE

3947	<G,>	<U0122>	LATIN CAPITAL LETTER G WITH CEDILLA
3948	<g,>	<U0123>	LATIN SMALL LETTER G WITH CEDILLA
3949	<H/>	<U0124>	LATIN CAPITAL LETTER H WITH CIRCUMFLEX
3950	<h/>	<U0125>	LATIN SMALL LETTER H WITH CIRCUMFLEX
3951	<H//>	<U0126>	LATIN CAPITAL LETTER H WITH STROKE
3952	<h//>	<U0127>	LATIN SMALL LETTER H WITH STROKE
3953	<I?>	<U0128>	LATIN CAPITAL LETTER I WITH TILDE
3954	<i?>	<U0129>	LATIN SMALL LETTER I WITH TILDE
3955	<I->	<U012A>	LATIN CAPITAL LETTER I WITH MACRON
3956	<i->	<U012B>	LATIN SMALL LETTER I WITH MACRON
3957	<I(>	<U012C>	LATIN CAPITAL LETTER I WITH BREVE
3958	<i(>	<U012D>	LATIN SMALL LETTER I WITH BREVE
3959	<I;>	<U012E>	LATIN CAPITAL LETTER I WITH OGONEK
3960	<i;>	<U012F>	LATIN SMALL LETTER I WITH OGONEK
3961	<I.>	<U0130>	LATIN CAPITAL LETTER I WITH DOT ABOVE
3962	<i.>	<U0131>	LATIN SMALL LETTER DOTLESS I
3963	<IJ>	<U0132>	LATIN CAPITAL LIGATURE IJ
3964	<i;j>	<U0133>	LATIN SMALL LIGATURE IJ
3965	<J/>	<U0134>	LATIN CAPITAL LETTER J WITH CIRCUMFLEX
3966	<j/>	<U0135>	LATIN SMALL LETTER J WITH CIRCUMFLEX
3967	<K,>	<U0136>	LATIN CAPITAL LETTER K WITH CEDILLA
3968	<k,>	<U0137>	LATIN SMALL LETTER K WITH CEDILLA
3969	<kK>	<U0138>	LATIN SMALL LETTER KRA (Greenlandic)
3970	<L'>	<U0139>	LATIN CAPITAL LETTER L WITH ACUTE
3971	<l'>	<U013A>	LATIN SMALL LETTER L WITH ACUTE
3972	<L,>	<U013B>	LATIN CAPITAL LETTER L WITH CEDILLA
3973	<l,>	<U013C>	LATIN SMALL LETTER L WITH CEDILLA
3974	<L<>	<U013D>	LATIN CAPITAL LETTER L WITH CARON
3975	<l<>	<U013E>	LATIN SMALL LETTER L WITH CARON
3976	<L.>	<U013F>	LATIN CAPITAL LETTER L WITH MIDDLE DOT
3977	<l.>	<U0140>	LATIN SMALL LETTER L WITH MIDDLE DOT
3978	<L//>	<U0141>	LATIN CAPITAL LETTER L WITH STROKE
3979	<l//>	<U0142>	LATIN SMALL LETTER L WITH STROKE
3980	<N'>	<U0143>	LATIN CAPITAL LETTER N WITH ACUTE
3981	<n'>	<U0144>	LATIN SMALL LETTER N WITH ACUTE
3982	<N,>	<U0145>	LATIN CAPITAL LETTER N WITH CEDILLA
3983	<n,>	<U0146>	LATIN SMALL LETTER N WITH CEDILLA
3984	<N<>	<U0147>	LATIN CAPITAL LETTER N WITH CARON
3985	<n<>	<U0148>	LATIN SMALL LETTER N WITH CARON
3986	<'n>	<U0149>	LATIN SMALL LETTER N PRECEDED BY APOSTROPHE
3987	<NG>	<U014A>	LATIN CAPITAL LETTER ENG (Sami)
3988	<ng>	<U014B>	LATIN SMALL LETTER ENG (Sami)
3989	<O->	<U014C>	LATIN CAPITAL LETTER O WITH MACRON
3990	<o->	<U014D>	LATIN SMALL LETTER O WITH MACRON
3991	<O(>	<U014E>	LATIN CAPITAL LETTER O WITH BREVE
3992	<o(>	<U014F>	LATIN SMALL LETTER O WITH BREVE
3993	<O">	<U0150>	LATIN CAPITAL LETTER O WITH DOUBLE ACUTE
3994	<o">	<U0151>	LATIN SMALL LETTER O WITH DOUBLE ACUTE
3995	<OE>	<U0152>	LATIN CAPITAL LIGATURE OE
3996	<oe>	<U0153>	LATIN SMALL LIGATURE OE
3997	<R'>	<U0154>	LATIN CAPITAL LETTER R WITH ACUTE
3998	<r'>	<U0155>	LATIN SMALL LETTER R WITH ACUTE
3999	<R,>	<U0156>	LATIN CAPITAL LETTER R WITH CEDILLA
4000	<r,>	<U0157>	LATIN SMALL LETTER R WITH CEDILLA
4001	<R<>	<U0158>	LATIN CAPITAL LETTER R WITH CARON
4002	<r<>	<U0159>	LATIN SMALL LETTER R WITH CARON
4003	<S'>	<U015A>	LATIN CAPITAL LETTER S WITH ACUTE
4004	<s'>	<U015B>	LATIN SMALL LETTER S WITH ACUTE
4005	<S//>	<U015C>	LATIN CAPITAL LETTER S WITH CIRCUMFLEX
4006	<s//>	<U015D>	LATIN SMALL LETTER S WITH CIRCUMFLEX
4007	<S,>	<U015E>	LATIN CAPITAL LETTER S WITH CEDILLA
4008	<s,>	<U015F>	LATIN SMALL LETTER S WITH CEDILLA
4009	<S<>	<U0160>	LATIN CAPITAL LETTER S WITH CARON
4010	<s<>	<U0161>	LATIN SMALL LETTER S WITH CARON
4011	<T,>	<U0162>	LATIN CAPITAL LETTER T WITH CEDILLA
4012	<t,>	<U0163>	LATIN SMALL LETTER T WITH CEDILLA
4013	<T<>	<U0164>	LATIN CAPITAL LETTER T WITH CARON
4014	<t<>	<U0165>	LATIN SMALL LETTER T WITH CARON
4015	<T//>	<U0166>	LATIN CAPITAL LETTER T WITH STROKE
4016	<t//>	<U0167>	LATIN SMALL LETTER T WITH STROKE
4017	<U?>	<U0168>	LATIN CAPITAL LETTER U WITH TILDE
4018	<u?>	<U0169>	LATIN SMALL LETTER U WITH TILDE
4019	<U->	<U016A>	LATIN CAPITAL LETTER U WITH MACRON
4020	<u->	<U016B>	LATIN SMALL LETTER U WITH MACRON
4021	<U(>	<U016C>	LATIN CAPITAL LETTER U WITH BREVE
4022	<u(>	<U016D>	LATIN SMALL LETTER U WITH BREVE
4023	<U0>	<U016E>	LATIN CAPITAL LETTER U WITH RING ABOVE
4024	<u0>	<U016F>	LATIN SMALL LETTER U WITH RING ABOVE
4025	<U">	<U0170>	LATIN CAPITAL LETTER U WITH DOUBLE ACUTE
4026	<u">	<U0171>	LATIN SMALL LETTER U WITH DOUBLE ACUTE
4027	<U;>	<U0172>	LATIN CAPITAL LETTER U WITH OGONEK
4028	<u;i>	<U0173>	LATIN SMALL LETTER U WITH OGONEK
4029	<W//>	<U0174>	LATIN CAPITAL LETTER W WITH CIRCUMFLEX
4030	<w//>	<U0175>	LATIN SMALL LETTER W WITH CIRCUMFLEX
4031	<Y//>	<U0176>	LATIN CAPITAL LETTER Y WITH CIRCUMFLEX
4032	<y//>	<U0177>	LATIN SMALL LETTER Y WITH CIRCUMFLEX
4033	<Y:>	<U0178>	LATIN CAPITAL LETTER Y WITH DIAERESIS
4034	<z'>	<U0179>	LATIN CAPITAL LETTER Z WITH ACUTE

4035	<z'>	<U017A>	LATIN SMALL LETTER Z WITH ACUTE
4036	<Z.>	<U017B>	LATIN CAPITAL LETTER Z WITH DOT ABOVE
4037	<z.>	<U017C>	LATIN SMALL LETTER Z WITH DOT ABOVE
4038	<Z<>	<U017D>	LATIN CAPITAL LETTER Z WITH CARON
4039	<z<>	<U017E>	LATIN SMALL LETTER Z WITH CARON
4040	<s1>	<U017F>	LATIN SMALL LETTER LONG S
4041	<b//>	<U0180>	LATIN SMALL LETTER B WITH STROKE
4042	<B2>	<U0181>	LATIN CAPITAL LETTER B WITH HOOK
4043	<C2>	<U0187>	LATIN CAPITAL LETTER C WITH HOOK
4044	<c2>	<U0188>	LATIN SMALL LETTER C WITH HOOK
4045	<F2>	<U0191>	LATIN CAPITAL LETTER F WITH HOOK
4046	<f2>	<U0192>	LATIN SMALL LETTER F WITH HOOK
4047	<K2>	<U0198>	LATIN CAPITAL LETTER K WITH HOOK
4048	<k2>	<U0199>	LATIN SMALL LETTER K WITH HOOK
4049	<O9>	<U01A0>	LATIN CAPITAL LETTER O WITH HORN
4050	<o9>	<U01A1>	LATIN SMALL LETTER O WITH HORN
4051	<OI>	<U01A2>	LATIN CAPITAL LETTER OI
4052	<o1>	<U01A3>	LATIN SMALL LETTER OI
4053	<yr>	<U01A6>	LATIN LETTER YR
4054	<U9>	<U01AF>	LATIN CAPITAL LETTER U WITH HORN
4055	<u9>	<U01B0>	LATIN SMALL LETTER U WITH HORN
4056	<Z//>	<U01B5>	LATIN CAPITAL LETTER Z WITH STROKE
4057	<z//>	<U01B6>	LATIN SMALL LETTER Z WITH STROKE
4058	<ED>	<U01B7>	LATIN CAPITAL LETTER EZH
4059	<DZ<>	<U01C4>	LATIN CAPITAL LETTER DZ WITH CARON
4060	<Dz<>	<U01C5>	LATIN CAPITAL LETTER D WITH SMALL LETTER Z WITH CARON
4061	<dz<>	<U01C6>	LATIN SMALL LETTER DZ WITH CARON
4062	<LJ3>	<U01C7>	LATIN CAPITAL LETTER LJ
4063	<Lj3>	<U01C8>	LATIN CAPITAL LETTER L WITH SMALL LETTER J
4064	<lj3>	<U01C9>	LATIN SMALL LETTER LJ
4065	<NJ3>	<U01CA>	LATIN CAPITAL LETTER NJ
4066	<Nj3>	<U01CB>	LATIN CAPITAL LETTER N WITH SMALL LETTER J
4067	<nj3>	<U01CC>	LATIN SMALL LETTER NJ
4068	<A<>	<U01CD>	LATIN CAPITAL LETTER A WITH CARON
4069	<a<>	<U01CE>	LATIN SMALL LETTER A WITH CARON
4070	<I<>	<U01CF>	LATIN CAPITAL LETTER I WITH CARON
4071	<i<>	<U01D0>	LATIN SMALL LETTER I WITH CARON
4072	<O<>	<U01D1>	LATIN CAPITAL LETTER O WITH CARON
4073	<o<>	<U01D2>	LATIN SMALL LETTER O WITH CARON
4074	<U<>	<U01D3>	LATIN CAPITAL LETTER U WITH CARON
4075	<u<>	<U01D4>	LATIN SMALL LETTER U WITH CARON
4076	<U:->	<U01D5>	LATIN CAPITAL LETTER U WITH DIAERESIS AND MACRON
4077	<u:->	<U01D6>	LATIN SMALL LETTER U WITH DIAERESIS AND MACRON
4078	<U:'>	<U01D7>	LATIN CAPITAL LETTER U WITH DIAERESIS AND ACUTE
4079	<u:'>	<U01D8>	LATIN SMALL LETTER U WITH DIAERESIS AND ACUTE
4080	<U:<>	<U01D9>	LATIN CAPITAL LETTER U WITH DIAERESIS AND CARON
4081	<u:<>	<U01DA>	LATIN SMALL LETTER U WITH DIAERESIS AND CARON
4082	<U:!:>	<U01DB>	LATIN CAPITAL LETTER U WITH DIAERESIS AND GRAVE
4083	<u:!:>	<U01DC>	LATIN SMALL LETTER U WITH DIAERESIS AND GRAVE
4084	<e1>	<U01DD>	LATIN SMALL LETTER TURNED E
4085	<A1>	<U01DE>	LATIN CAPITAL LETTER A WITH DIAERESIS AND MACRON
4086	<a1>	<U01DF>	LATIN SMALL LETTER A WITH DIAERESIS AND MACRON
4087	<A7>	<U01E0>	LATIN CAPITAL LETTER A WITH DOT ABOVE AND MACRON
4088	<a7>	<U01E1>	LATIN SMALL LETTER A WITH DOT ABOVE AND MACRON
4089	<A3>	<U01E2>	LATIN CAPITAL LETTER AE WITH MACRON (ash)
4090	<a3>	<U01E3>	LATIN SMALL LETTER AE WITH MACRON (ash)
4091	<G//>	<U01E4>	LATIN CAPITAL LETTER G WITH STROKE
4092	<g//>	<U01E5>	LATIN SMALL LETTER G WITH STROKE
4093	<G<>	<U01E6>	LATIN CAPITAL LETTER G WITH CARON
4094	<g<>	<U01E7>	LATIN SMALL LETTER G WITH CARON
4095	<K<>	<U01E8>	LATIN CAPITAL LETTER K WITH CARON
4096	<k<>	<U01E9>	LATIN SMALL LETTER K WITH CARON
4097	<O;>	<U01EA>	LATIN CAPITAL LETTER O WITH OGONEK
4098	<o;>	<U01EB>	LATIN SMALL LETTER O WITH OGONEK
4099	<O1>	<U01EC>	LATIN CAPITAL LETTER O WITH OGONEK AND MACRON
4100	<o1>	<U01ED>	LATIN SMALL LETTER O WITH OGONEK AND MACRON
4101	<EZ>	<U01EE>	LATIN CAPITAL LETTER EZH WITH CARON
4102	<ez>	<U01EF>	LATIN SMALL LETTER EZH WITH CARON
4103	<j<>	<U01F0>	LATIN SMALL LETTER J WITH CARON
4104	<DZ3>	<U01F1>	LATIN CAPITAL LETTER DZ
4105	<Dz3>	<U01F2>	LATIN CAPITAL LETTER D WITH SMALL LETTER Z
4106	<d3>	<U01F3>	LATIN SMALL LETTER DZ
4107	<G'>	<U01F4>	LATIN CAPITAL LETTER G WITH ACUTE
4108	<g'>	<U01F5>	LATIN SMALL LETTER G WITH ACUTE
4109	<AA'>	<U01FA>	LATIN CAPITAL LETTER A WITH RING ABOVE AND ACUTE
4110	<aa'>	<U01FB>	LATIN SMALL LETTER A WITH RING ABOVE AND ACUTE
4111	<AE'>	<U01FC>	LATIN CAPITAL LETTER AE WITH ACUTE (ash)
4112	<ae'>	<U01FD>	LATIN SMALL LETTER AE WITH ACUTE (ash)
4113	<O//>	<U01FE>	LATIN CAPITAL LETTER O WITH STROKE AND ACUTE
4114	<o//>	<U01FF>	LATIN SMALL LETTER O WITH STROKE AND ACUTE
4115	<A!!>	<U0200>	LATIN CAPITAL LETTER A WITH DOUBLE GRAVE
4116	<a!!>	<U0201>	LATIN SMALL LETTER A WITH DOUBLE GRAVE
4117	<A>	<U0202>	LATIN CAPITAL LETTER A WITH INVERTED BREVE
4118	<a>	<U0203>	LATIN SMALL LETTER A WITH INVERTED BREVE
4119	<E!!>	<U0204>	LATIN CAPITAL LETTER E WITH DOUBLE GRAVE
4120	<e!!>	<U0205>	LATIN SMALL LETTER E WITH DOUBLE GRAVE
4121	<E>	<U0206>	LATIN CAPITAL LETTER E WITH INVERTED BREVE
4122	<e>	<U0207>	LATIN SMALL LETTER E WITH INVERTED BREVE

4 23	<I!!>	<U0208>	LATIN CAPITAL LETTER I WITH DOUBLE GRAVE
4 24	<i!!>	<U0209>	LATIN SMALL LETTER I WITH DOUBLE GRAVE
4 25	<I>	<U020A>	LATIN CAPITAL LETTER I WITH INVERTED BREVE
4 26	<i>	<U020B>	LATIN SMALL LETTER I WITH INVERTED BREVE
4 27	<O!!>	<U020C>	LATIN CAPITAL LETTER O WITH DOUBLE GRAVE
4 28	<o!!>	<U020D>	LATIN SMALL LETTER O WITH DOUBLE GRAVE
4 29	<O>	<U020E>	LATIN CAPITAL LETTER O WITH INVERTED BREVE
4 30	<o>	<U020F>	LATIN SMALL LETTER O WITH INVERTED BREVE
4 31	<R!!>	<U0210>	LATIN CAPITAL LETTER R WITH DOUBLE GRAVE
4 32	<r!!>	<U0211>	LATIN SMALL LETTER R WITH DOUBLE GRAVE
4 33	<R>	<U0212>	LATIN CAPITAL LETTER R WITH INVERTED BREVE
4 34	<r>	<U0213>	LATIN SMALL LETTER R WITH INVERTED BREVE
4 35	<U!!>	<U0214>	LATIN CAPITAL LETTER U WITH DOUBLE GRAVE
4 36	<u!!>	<U0215>	LATIN SMALL LETTER U WITH DOUBLE GRAVE
4 37	<U>	<U0216>	LATIN CAPITAL LETTER U WITH INVERTED BREVE
4 38	<u>	<U0217>	LATIN SMALL LETTER U WITH INVERTED BREVE
4 39	<r1>	<U027C>	LATIN SMALL LETTER R WITH LONG LEG
4 40	<ed>	<U0292>	LATIN SMALL LETTER EZH
4 41	<;S>	<U02BB>	MODIFIER LETTER TURNED COMMA
4 42	</>>	<U02C6>	MODIFIER LETTER CIRCUMFLEX ACCENT
4 43	<'>>	<U02C7>	CARON (Mandarin Chinese third tone)
4 44	<1->	<U02C9>	MODIFIER LETTER MACRON (Mandarin Chinese first tone)
4 45	<1!>	<U02CB>	MODIFIER LETTER GRAVE ACCENT (Mandarin Chinese fourth tone)
4 46	<'(>	<U02D8>	BREVE
4 47	<'.>	<U02D9>	DOT ABOVE (Mandarin Chinese light tone)
4 48	<'0>	<U02DA>	RING ABOVE
4 49	<'1>	<U02DB>	OGONEK
4 50	<1?>	<U02DC>	SMALL TILDE
4 51	<'">	<U02DD>	DOUBLE ACUTE ACCENT
4 52	<'G>	<U0374>	GREEK NUMERAL SIGN (Dexia keraia)
4 53	<,G>	<U0375>	GREEK LOWER NUMERAL SIGN (Aristeri keraia)
4 54	<j3>	<U037A>	GREEK YPOGEGRAMMENI
4 55	<?%>	<U037E>	GREEK QUESTION MARK (Erotimatiiko)
4 56	<'*>	<U0384>	GREEK TONOS
4 57	<'%>	<U0385>	GREEK DIALYTIKA TONOS
4 58	<A%>	<U0386>	GREEK CAPITAL LETTER ALPHA WITH TONOS
4 59	<. *>	<U0387>	GREEK ANO TELEIA
4 60	<E%>	<U0388>	GREEK CAPITAL LETTER EPSILON WITH TONOS
4 61	<Y%>	<U0389>	GREEK CAPITAL LETTER ETA WITH TONOS
4 62	<I%>	<U038A>	GREEK CAPITAL LETTER IOTA WITH TONOS
4 63	<O%>	<U038C>	GREEK CAPITAL LETTER OMICRON WITH TONOS
4 64	<U%>	<U038E>	GREEK CAPITAL LETTER UPSILON WITH TONOS
4 65	<W%>	<U038F>	GREEK CAPITAL LETTER OMEGA WITH TONOS
4 66	<i3>	<U0390>	GREEK SMALL LETTER IOTA WITH DIALYTIKA AND TONOS
4 67	<A*>	<U0391>	GREEK CAPITAL LETTER ALPHA
4 68	<B*>	<U0392>	GREEK CAPITAL LETTER BETA
4 69	<G*>	<U0393>	GREEK CAPITAL LETTER GAMMA
4 70	<D*>	<U0394>	GREEK CAPITAL LETTER DELTA
4 71	<E*>	<U0395>	GREEK CAPITAL LETTER EPSILON
4 72	<Z*>	<U0396>	GREEK CAPITAL LETTER ZETA
4 73	<Y*>	<U0397>	GREEK CAPITAL LETTER ETA
4 74	<H*>	<U0398>	GREEK CAPITAL LETTER THETA
4 75	<I*>	<U0399>	GREEK CAPITAL LETTER IOTA
4 76	<K*>	<U039A>	GREEK CAPITAL LETTER KAPPA
4 77	<L*>	<U039B>	GREEK CAPITAL LETTER LAMDA
4 78	<M*>	<U039C>	GREEK CAPITAL LETTER MU
4 79	<N*>	<U039D>	GREEK CAPITAL LETTER NU
4 80	<C*>	<U039E>	GREEK CAPITAL LETTER XI
4 81	<O*>	<U039F>	GREEK CAPITAL LETTER OMICRON
4 82	<P*>	<U03A0>	GREEK CAPITAL LETTER PI
4 83	<R*>	<U03A1>	GREEK CAPITAL LETTER RHO
4 84	<S*>	<U03A3>	GREEK CAPITAL LETTER SIGMA
4 85	<T*>	<U03A4>	GREEK CAPITAL LETTER TAU
4 86	<U*>	<U03A5>	GREEK CAPITAL LETTER UPSILON
4 87	<F*>	<U03A6>	GREEK CAPITAL LETTER PHI
4 88	<X*>	<U03A7>	GREEK CAPITAL LETTER CHI
4 89	<Q*>	<U03A8>	GREEK CAPITAL LETTER PSI
4 90	<W*>	<U03A9>	GREEK CAPITAL LETTER OMEGA
4 91	<J*>	<U03AA>	GREEK CAPITAL LETTER IOTA WITH DIALYTIKA
4 92	<V*>	<U03AB>	GREEK CAPITAL LETTER UPSILON WITH DIALYTIKA
4 93	<a%>	<U03AC>	GREEK SMALL LETTER ALPHA WITH TONOS
4 94	<e%>	<U03AD>	GREEK SMALL LETTER EPSILON WITH TONOS
4 95	<y%>	<U03AE>	GREEK SMALL LETTER ETA WITH TONOS
4 96	<i%>	<U03AF>	GREEK SMALL LETTER IOTA WITH TONOS
4 97	<u3>	<U03B0>	GREEK SMALL LETTER UPSILON WITH DIALYTIKA AND TONOS
4 98	<a*>	<U03B1>	GREEK SMALL LETTER ALPHA
4 99	<b*>	<U03B2>	GREEK SMALL LETTER BETA
4 00	<g*>	<U03B3>	GREEK SMALL LETTER GAMMA
4 01	<d*>	<U03B4>	GREEK SMALL LETTER DELTA
4 02	<e*>	<U03B5>	GREEK SMALL LETTER EPSILON
4 03	<z*>	<U03B6>	GREEK SMALL LETTER ZETA
4 04	<y*>	<U03B7>	GREEK SMALL LETTER ETA
4 05	<h*>	<U03B8>	GREEK SMALL LETTER THETA
4 06	<i*>	<U03B9>	GREEK SMALL LETTER IOTA
4 07	<k*>	<U03BA>	GREEK SMALL LETTER KAPPA
4 08	<l*>	<U03BB>	GREEK SMALL LETTER LAMDA
4 09	<m*>	<U03BC>	GREEK SMALL LETTER MU
4 10	<n*>	<U03BD>	GREEK SMALL LETTER NU

4211	<c*>	<U03BE>	GREEK SMALL LETTER XI
4212	<o*>	<U03BF>	GREEK SMALL LETTER OMICRON
4213	<p*>	<U03C0>	GREEK SMALL LETTER PI
4214	<r*>	<U03C1>	GREEK SMALL LETTER RHO
4215	<*s>	<U03C2>	GREEK SMALL LETTER FINAL SIGMA
4216	<s*>	<U03C3>	GREEK SMALL LETTER SIGMA
4217	<t*>	<U03C4>	GREEK SMALL LETTER TAU
4218	<u*>	<U03C5>	GREEK SMALL LETTER UPSILON
4219	<f*>	<U03C6>	GREEK SMALL LETTER PHI
4220	<x*>	<U03C7>	GREEK SMALL LETTER CHI
4221	<q*>	<U03C8>	GREEK SMALL LETTER PSI
4222	<w*>	<U03C9>	GREEK SMALL LETTER OMEGA
4223	<j*>	<U03CA>	GREEK SMALL LETTER IOTA WITH DIALYTIKA
4224	<v*>	<U03CB>	GREEK SMALL LETTER UPSILON WITH DIALYTIKA
4225	<o%>	<U03CC>	GREEK SMALL LETTER OMICRON WITH TONOS
4226	<u%>	<U03CD>	GREEK SMALL LETTER UPSILON WITH TONOS
4227	<w%>	<U03CE>	GREEK SMALL LETTER OMEGA WITH TONOS
4228	<b3>	<U03D0>	GREEK BETA SYMBOL
4229	<T3>	<U03DA>	GREEK LETTER STIGMA
4230	<M3>	<U03DC>	GREEK LETTER DIGAMMA
4231	<K3>	<U03DE>	GREEK LETTER KOPPA
4232	<P3>	<U03E0>	GREEK LETTER SAMPI
4233	<IO>	<U0401>	CYRILLIC CAPITAL LETTER IO
4234	<D%>	<U0402>	CYRILLIC CAPITAL LETTER DJE (Serbocroatian)
4235	<G%>	<U0403>	CYRILLIC CAPITAL LETTER GJE
4236	<IE>	<U0404>	CYRILLIC CAPITAL LETTER UKRAINIAN IE
4237	<DS>	<U0405>	CYRILLIC CAPITAL LETTER DZE
4238	<II>	<U0406>	CYRILLIC CAPITAL LETTER BYELORUSSIAN-UKRAINIAN I
4239	<YI>	<U0407>	CYRILLIC CAPITAL LETTER YI (Ukrainian)
4240	<J%>	<U0408>	CYRILLIC CAPITAL LETTER JE
4241	<LJ>	<U0409>	CYRILLIC CAPITAL LETTER LJE
4242	<NJ>	<U040A>	CYRILLIC CAPITAL LETTER NJE
4243	<Ts>	<U040B>	CYRILLIC CAPITAL LETTER TSHE (Serbocroatian)
4244	<KJ>	<U040C>	CYRILLIC CAPITAL LETTER KJE
4245	<V%>	<U040E>	CYRILLIC CAPITAL LETTER SHORT U (Byelorussian)
4246	<DZ>	<U040F>	CYRILLIC CAPITAL LETTER DZHE
4247	<A=>	<U0410>	CYRILLIC CAPITAL LETTER A
4248	<B=>	<U0411>	CYRILLIC CAPITAL LETTER BE
4249	<V=>	<U0412>	CYRILLIC CAPITAL LETTER VE
4250	<G=>	<U0413>	CYRILLIC CAPITAL LETTER GHE
4251	<D=>	<U0414>	CYRILLIC CAPITAL LETTER DE
4252	<E=>	<U0415>	CYRILLIC CAPITAL LETTER IE
4253	<Z%>	<U0416>	CYRILLIC CAPITAL LETTER ZHE
4254	<Z=>	<U0417>	CYRILLIC CAPITAL LETTER ZE
4255	<I=>	<U0418>	CYRILLIC CAPITAL LETTER I
4256	<J=>	<U0419>	CYRILLIC CAPITAL LETTER SHORT I
4257	<K=>	<U041A>	CYRILLIC CAPITAL LETTER KA
4258	<L=>	<U041B>	CYRILLIC CAPITAL LETTER EL
4259	<M=>	<U041C>	CYRILLIC CAPITAL LETTER EM
4260	<N=>	<U041D>	CYRILLIC CAPITAL LETTER EN
4261	<O=>	<U041E>	CYRILLIC CAPITAL LETTER O
4262	<P=>	<U041F>	CYRILLIC CAPITAL LETTER PE
4263	<R=>	<U0420>	CYRILLIC CAPITAL LETTER ER
4264	<S=>	<U0421>	CYRILLIC CAPITAL LETTER ES
4265	<T=>	<U0422>	CYRILLIC CAPITAL LETTER TE
4266	<U=>	<U0423>	CYRILLIC CAPITAL LETTER U
4267	<F=>	<U0424>	CYRILLIC CAPITAL LETTER EF
4268	<H=>	<U0425>	CYRILLIC CAPITAL LETTER HA
4269	<C=>	<U0426>	CYRILLIC CAPITAL LETTER TSE
4270	<C%>	<U0427>	CYRILLIC CAPITAL LETTER CHE
4271	<S%>	<U0428>	CYRILLIC CAPITAL LETTER SHA
4272	<SC>	<U0429>	CYRILLIC CAPITAL LETTER SHCHA
4273	<=*>	<U042A>	CYRILLIC CAPITAL LETTER HARD SIGN
4274	<Y=>	<U042B>	CYRILLIC CAPITAL LETTER YERU
4275	<%>	<U042C>	CYRILLIC CAPITAL LETTER SOFT SIGN
4276	<JE>	<U042D>	CYRILLIC CAPITAL LETTER E
4277	<JU>	<U042E>	CYRILLIC CAPITAL LETTER YU
4278	<JA>	<U042F>	CYRILLIC CAPITAL LETTER YA
4279	<a=>	<U0430>	CYRILLIC SMALL LETTER A
4280	<b=>	<U0431>	CYRILLIC SMALL LETTER BE
4281	<v=>	<U0432>	CYRILLIC SMALL LETTER VE
4282	<g=>	<U0433>	CYRILLIC SMALL LETTER GHE
4283	<d=>	<U0434>	CYRILLIC SMALL LETTER DE
4284	<e=>	<U0435>	CYRILLIC SMALL LETTER IE
4285	<z%>	<U0436>	CYRILLIC SMALL LETTER ZHE
4286	<z=>	<U0437>	CYRILLIC SMALL LETTER ZE
4287	<i=>	<U0438>	CYRILLIC SMALL LETTER I
4288	<j=>	<U0439>	CYRILLIC SMALL LETTER SHORT I
4289	<k=>	<U043A>	CYRILLIC SMALL LETTER KA
4290	<l=>	<U043B>	CYRILLIC SMALL LETTER EL
4291	<m=>	<U043C>	CYRILLIC SMALL LETTER EM
4292	<n=>	<U043D>	CYRILLIC SMALL LETTER EN
4293	<o=>	<U043E>	CYRILLIC SMALL LETTER O
4294	<p=>	<U043F>	CYRILLIC SMALL LETTER PE
4295	<r=>	<U0440>	CYRILLIC SMALL LETTER ER
4296	<s=>	<U0441>	CYRILLIC SMALL LETTER ES
4297	<t=>	<U0442>	CYRILLIC SMALL LETTER TE
4298	<u=>	<U0443>	CYRILLIC SMALL LETTER U

4299	<f=>	<U0444>	CYRILLIC SMALL LETTER EF
4300	<h=>	<U0445>	CYRILLIC SMALL LETTER HA
4301	<c=>	<U0446>	CYRILLIC SMALL LETTER TSE
4302	<c%>	<U0447>	CYRILLIC SMALL LETTER CHE
4303	<s%>	<U0448>	CYRILLIC SMALL LETTER SHA
4304	<sc>	<U0449>	CYRILLIC SMALL LETTER SHCHA
4305	<= '>	<U044A>	CYRILLIC SMALL LETTER HARD SIGN
4306	<y=>	<U044B>	CYRILLIC SMALL LETTER YERU
4307	<% '>	<U044C>	CYRILLIC SMALL LETTER SOFT SIGN
4308	<je>	<U044D>	CYRILLIC SMALL LETTER E
4309	<ju>	<U044E>	CYRILLIC SMALL LETTER YU
4310	<ja>	<U044F>	CYRILLIC SMALL LETTER YA
4311	<io>	<U0451>	CYRILLIC SMALL LETTER IO
4312	<d%>	<U0452>	CYRILLIC SMALL LETTER DJE (SerboCroatian)
4313	<g%>	<U0453>	CYRILLIC SMALL LETTER GJE
4314	<ie>	<U0454>	CYRILLIC SMALL LETTER UKRAINIAN IE
4315	<ds>	<U0455>	CYRILLIC SMALL LETTER DZE
4316	<i>	<U0456>	CYRILLIC SMALL LETTER BYELORUSSIAN-UKRAINIAN I
4317	<yi>	<U0457>	CYRILLIC SMALL LETTER YI (Ukrainian)
4318	<j%>	<U0458>	CYRILLIC SMALL LETTER JE
4319	<l>	<U0459>	CYRILLIC SMALL LETTER LJE
4320	<n>	<U045A>	CYRILLIC SMALL LETTER NJE
4321	<ts>	<U045B>	CYRILLIC SMALL LETTER TSHE (SerboCroatian)
4322	<k>	<U045C>	CYRILLIC SMALL LETTER KJE
4323	<v%>	<U045E>	CYRILLIC SMALL LETTER SHORT U (Byelorussian)
4324	<dz>	<U045F>	CYRILLIC SMALL LETTER DZHE
4325	<Y>	<U0462>	CYRILLIC CAPITAL LETTER YAT
4326	<y>	<U0463>	CYRILLIC SMALL LETTER YAT
4327	<O>	<U046A>	CYRILLIC CAPITAL LETTER BIG YUS
4328	<o>	<U046B>	CYRILLIC SMALL LETTER BIG YUS
4329	<F>	<U0472>	CYRILLIC CAPITAL LETTER FITA
4330	<f>	<U0473>	CYRILLIC SMALL LETTER FITA
4331	<V>	<U0474>	CYRILLIC CAPITAL LETTER IZHITSA
4332	<v>	<U0475>	CYRILLIC SMALL LETTER IZHITSA
4333	<C>	<U0480>	CYRILLIC CAPITAL LETTER KOPPA
4334	<c>	<U0481>	CYRILLIC SMALL LETTER KOPPA
4335	<G>	<U0490>	CYRILLIC CAPITAL LETTER GHE WITH UPTURN
4336	<g>	<U0491>	CYRILLIC SMALL LETTER GHE WITH UPTURN
4337	<A>	<U05D0>	HEBREW LETTER ALEF
4338	<B>	<U05D1>	HEBREW LETTER BET
4339	<G>	<U05D2>	HEBREW LETTER GIMEL
4340	<D>	<U05D3>	HEBREW LETTER DALET
4341	<H>	<U05D4>	HEBREW LETTER HE
4342	<W>	<U05D5>	HEBREW LETTER VAV
4343	<Z>	<U05D6>	HEBREW LETTER ZAYIN
4344	<X>	<U05D7>	HEBREW LETTER HET
4345	<T>	<U05D8>	HEBREW LETTER TET
4346	<J>	<U05D9>	HEBREW LETTER YOD
4347	<K%>	<U05DA>	HEBREW LETTER FINAL KAF
4348	<K>	<U05DB>	HEBREW LETTER KAF
4349	<L>	<U05DC>	HEBREW LETTER LAMED
4350	<M%>	<U05DD>	HEBREW LETTER FINAL MEM
4351	<M>	<U05DE>	HEBREW LETTER MEM
4352	<N%>	<U05DF>	HEBREW LETTER FINAL NUN
4353	<N>	<U05E0>	HEBREW LETTER NUN
4354	<S>	<U05E1>	HEBREW LETTER SAMEKH
4355	<E>	<U05E2>	HEBREW LETTER AYIN
4356	<P%>	<U05E3>	HEBREW LETTER FINAL PE
4357	<P>	<U05E4>	HEBREW LETTER PE
4358	<Z>	<U05E5>	HEBREW LETTER FINAL TSADI
4359	<Z>	<U05E6>	HEBREW LETTER TSADI
4360	<Q>	<U05E7>	HEBREW LETTER QOF
4361	<R>	<U05E8>	HEBREW LETTER RESH
4362	<Sh>	<U05E9>	HEBREW LETTER SHIN
4363	<T>	<U05EA>	HEBREW LETTER TAV
4364	<, +>	<U060C>	ARABIC COMMA
4365	<: +>	<U061B>	ARABIC SEMICOLON
4366	<? +>	<U061F>	ARABIC QUESTION MARK
4367	<H'>	<U0621>	ARABIC LETTER HAMZA
4368	<aM>	<U0622>	ARABIC LETTER ALEF WITH MADDA ABOVE
4369	<aH>	<U0623>	ARABIC LETTER ALEF WITH HAMZA ABOVE
4370	<wH>	<U0624>	ARABIC LETTER WAW WITH HAMZA ABOVE
4371	<ah>	<U0625>	ARABIC LETTER ALEF WITH HAMZA BELOW
4372	<yH>	<U0626>	ARABIC LETTER YEH WITH HAMZA ABOVE
4373	<a>	<U0627>	ARABIC LETTER ALEF
4374	<b>	<U0628>	ARABIC LETTER BEH
4375	<tm>	<U0629>	ARABIC LETTER TEH MARBUTA
4376	<t>	<U062A>	ARABIC LETTER TEH
4377	<tk>	<U062B>	ARABIC LETTER THEH
4378	<g>	<U062C>	ARABIC LETTER JEEEM
4379	<hk>	<U062D>	ARABIC LETTER HAH
4380	<x>	<U062E>	ARABIC LETTER KHAH
4381	<d>	<U062F>	ARABIC LETTER DAL
4382	<dk>	<U0630>	ARABIC LETTER THAL
4383	<r>	<U0631>	ARABIC LETTER REH
4384	<z>	<U0632>	ARABIC LETTER ZAIN
4385	<s>	<U0633>	ARABIC LETTER SEEN
4386	<sn>	<U0634>	ARABIC LETTER SHEEN

4387	<c+>	<U0635>	ARABIC LETTER SAD
4388	<d>	<U0636>	ARABIC LETTER DAD
4389	<t>	<U0637>	ARABIC LETTER TAH
4390	<zH>	<U0638>	ARABIC LETTER ZAH
4391	<e+>	<U0639>	ARABIC LETTER AIN
4392	<i+>	<U063A>	ARABIC LETTER GHAIN
4393	<+>	<U0640>	ARABIC TATWEEL
4394	<f+>	<U0641>	ARABIC LETTER FEH
4395	<q+>	<U0642>	ARABIC LETTER QAF
4396	<k+>	<U0643>	ARABIC LETTER KAF
4397	<l+>	<U0644>	ARABIC LETTER LAM
4398	<m+>	<U0645>	ARABIC LETTER MEEM
4399	<n+>	<U0646>	ARABIC LETTER NOON
4400	<h+>	<U0647>	ARABIC LETTER HEH
4401	<w+>	<U0648>	ARABIC LETTER WAW
4402	<j+>	<U0649>	ARABIC LETTER ALEF MAKSURA
4403	<y+>	<U064A>	ARABIC LETTER YEH
4404	<:+>	<U064B>	ARABIC FATHATAN
4405	<">	<U064C>	ARABIC DAMMATAN
4406	<=+>	<U064D>	ARABIC KASRATAN
4407	<//+>	<U064E>	ARABIC FATHA
4408	<'+>	<U064F>	ARABIC DAMMA
4409	<1+>	<U0650>	ARABIC KASRA
4410	<3+>	<U0651>	ARABIC SHADDA
4411	<0+>	<U0652>	ARABIC SUKUN
4412	<0a>	<U0660>	ARABIC-INDIC DIGIT ZERO
4413	<1a>	<U0661>	ARABIC-INDIC DIGIT ONE
4414	<2a>	<U0662>	ARABIC-INDIC DIGIT TWO
4415	<3a>	<U0663>	ARABIC-INDIC DIGIT THREE
4416	<4a>	<U0664>	ARABIC-INDIC DIGIT FOUR
4417	<5a>	<U0665>	ARABIC-INDIC DIGIT FIVE
4418	<6a>	<U0666>	ARABIC-INDIC DIGIT SIX
4419	<7a>	<U0667>	ARABIC-INDIC DIGIT SEVEN
4420	<8a>	<U0668>	ARABIC-INDIC DIGIT EIGHT
4421	<9a>	<U0669>	ARABIC-INDIC DIGIT NINE
4422	<aS>	<U0670>	ARABIC LETTER SUPERSCRIPT ALEF
4423	<p+>	<U067E>	ARABIC LETTER PEH
4424	<hH>	<U0681>	ARABIC LETTER HAH WITH HAMZA ABOVE
4425	<tC>	<U0686>	ARABIC LETTER TCHEH
4426	<zj>	<U0698>	ARABIC LETTER JEH
4427	<v+>	<U06A4>	ARABIC LETTER VEH
4428	<gf>	<U06AF>	ARABIC LETTER GAF
4429	<A-0>	<U1E00>	LATIN CAPITAL LETTER A WITH RING BELOW
4430	<a-0>	<U1E01>	LATIN SMALL LETTER A WITH RING BELOW
4431	<B,>	<U1E02>	LATIN CAPITAL LETTER B WITH DOT ABOVE
4432	<b,>	<U1E03>	LATIN SMALL LETTER B WITH DOT ABOVE
4433	<B,->	<U1E04>	LATIN CAPITAL LETTER B WITH DOT BELOW
4434	<b,->	<U1E05>	LATIN SMALL LETTER B WITH DOT BELOW
4435	<B,_>	<U1E06>	LATIN CAPITAL LETTER B WITH LINE BELOW
4436	<b,_>	<U1E07>	LATIN SMALL LETTER B WITH LINE BELOW
4437	<C,'>	<U1E08>	LATIN CAPITAL LETTER C WITH CEDILLA AND ACUTE
4438	<c,'>	<U1E09>	LATIN SMALL LETTER C WITH CEDILLA AND ACUTE
4439	<D,>	<U1E0A>	LATIN CAPITAL LETTER D WITH DOT ABOVE
4440	<d,>	<U1E0B>	LATIN SMALL LETTER D WITH DOT ABOVE
4441	<D-,>	<U1E0C>	LATIN CAPITAL LETTER D WITH DOT BELOW
4442	<d-,>	<U1E0D>	LATIN SMALL LETTER D WITH DOT BELOW
4443	<D,_>	<U1E0E>	LATIN CAPITAL LETTER D WITH LINE BELOW
4444	<d,_>	<U1E0F>	LATIN SMALL LETTER D WITH LINE BELOW
4445	<D,>	<U1E10>	LATIN CAPITAL LETTER D WITH CEDILLA
4446	<d,>	<U1E11>	LATIN SMALL LETTER D WITH CEDILLA
4447	<D-/>	<U1E12>	LATIN CAPITAL LETTER D WITH CIRCUMFLEX BELOW
4448	<d-/>	<U1E13>	LATIN SMALL LETTER D WITH CIRCUMFLEX BELOW
4449	<E-!>	<U1E14>	LATIN CAPITAL LETTER E WITH MACRON AND GRAVE
4450	<e-!>	<U1E15>	LATIN SMALL LETTER E WITH MACRON AND GRAVE
4451	<E-'/>	<U1E16>	LATIN CAPITAL LETTER E WITH MACRON AND ACUTE
4452	<e-'/>	<U1E17>	LATIN SMALL LETTER E WITH MACRON AND ACUTE
4453	<E-/>>	<U1E18>	LATIN CAPITAL LETTER E WITH CIRCUMFLEX BELOW
4454	<e-/>>	<U1E19>	LATIN SMALL LETTER E WITH CIRCUMFLEX BELOW
4455	<E-?>	<U1E1A>	LATIN CAPITAL LETTER E WITH TILDE BELOW
4456	<e-?>	<U1E1B>	LATIN SMALL LETTER E WITH TILDE BELOW
4457	<E,(>	<U1E1C>	LATIN CAPITAL LETTER E WITH CEDILLA AND BREVE
4458	<e,(>	<U1E1D>	LATIN SMALL LETTER E WITH CEDILLA AND BREVE
4459	<F,>	<U1E1E>	LATIN CAPITAL LETTER F WITH DOT ABOVE
4460	<f,>	<U1E1F>	LATIN SMALL LETTER F WITH DOT ABOVE
4461	<G->	<U1E20>	LATIN CAPITAL LETTER G WITH MACRON
4462	<g->	<U1E21>	LATIN SMALL LETTER G WITH MACRON
4463	<H,>	<U1E22>	LATIN CAPITAL LETTER H WITH DOT ABOVE
4464	<h,>	<U1E23>	LATIN SMALL LETTER H WITH DOT ABOVE
4465	<H,->	<U1E24>	LATIN CAPITAL LETTER H WITH DOT BELOW
4466	<h,->	<U1E25>	LATIN SMALL LETTER H WITH DOT BELOW
4467	<H,:>	<U1E26>	LATIN CAPITAL LETTER H WITH DIAERESIS
4468	<h,:>	<U1E27>	LATIN SMALL LETTER H WITH DIAERESIS
4469	<H,>	<U1E28>	LATIN CAPITAL LETTER H WITH CEDILLA
4470	<h,>	<U1E29>	LATIN SMALL LETTER H WITH CEDILLA
4471	<H-(>	<U1E2A>	LATIN CAPITAL LETTER H WITH BREVE BELOW
4472	<h-(>	<U1E2B>	LATIN SMALL LETTER H WITH BREVE BELOW
4473	<I-?>	<U1E2C>	LATIN CAPITAL LETTER I WITH TILDE BELOW
4474	<i-?>	<U1E2D>	LATIN SMALL LETTER I WITH TILDE BELOW

4475	<i:'>	<U1E2E>	LATIN CAPITAL LETTER I WITH DIAERESIS AND ACUTE
4476	<i:'>	<U1E2F>	LATIN SMALL LETTER I WITH DIAERESIS AND ACUTE
4477	<k:'>	<U1E30>	LATIN CAPITAL LETTER K WITH ACUTE
4478	<k:'>	<U1E31>	LATIN SMALL LETTER K WITH ACUTE
4479	<k-'>	<U1E32>	LATIN CAPITAL LETTER K WITH DOT BELOW
4480	<k-'>	<U1E33>	LATIN SMALL LETTER K WITH DOT BELOW
4481	<k_>	<U1E34>	LATIN CAPITAL LETTER K WITH LINE BELOW
4482	<k_>	<U1E35>	LATIN SMALL LETTER K WITH LINE BELOW
4483	<l-'>	<U1E36>	LATIN CAPITAL LETTER L WITH DOT BELOW
4484	<l-'>	<U1E37>	LATIN SMALL LETTER L WITH DOT BELOW
4485	<l--->	<U1E38>	LATIN CAPITAL LETTER L WITH DOT BELOW AND MACRON
4486	<l--->	<U1E39>	LATIN SMALL LETTER L WITH DOT BELOW AND MACRON
4487	<l_>	<U1E3A>	LATIN CAPITAL LETTER L WITH LINE BELOW
4488	<l_>	<U1E3B>	LATIN SMALL LETTER L WITH LINE BELOW
4489	<l-/>	<U1E3C>	LATIN CAPITAL LETTER L WITH CIRCUMFLEX BELOW
4490	<l-/>	<U1E3D>	LATIN SMALL LETTER L WITH CIRCUMFLEX BELOW
4491	<m:'>	<U1E3E>	LATIN CAPITAL LETTER M WITH ACUTE
4492	<m:'>	<U1E3F>	LATIN SMALL LETTER M WITH ACUTE
4493	<m:>	<U1E40>	LATIN CAPITAL LETTER M WITH DOT ABOVE
4494	<m:>	<U1E41>	LATIN SMALL LETTER M WITH DOT ABOVE
4495	<m-'>	<U1E42>	LATIN CAPITAL LETTER M WITH DOT BELOW
4496	<m-'>	<U1E43>	LATIN SMALL LETTER M WITH DOT BELOW
4497	<n:'>	<U1E44>	LATIN CAPITAL LETTER N WITH DOT ABOVE
4498	<n:>	<U1E45>	LATIN SMALL LETTER N WITH DOT ABOVE
4499	<n-'>	<U1E46>	LATIN CAPITAL LETTER N WITH DOT BELOW
4500	<n-'>	<U1E47>	LATIN SMALL LETTER N WITH DOT BELOW
4501	<n_>	<U1E48>	LATIN CAPITAL LETTER N WITH LINE BELOW
4502	<n_>	<U1E49>	LATIN SMALL LETTER N WITH LINE BELOW
4503	<n-/>	<U1E4A>	LATIN CAPITAL LETTER N WITH CIRCUMFLEX BELOW
4504	<n-/>	<U1E4B>	LATIN SMALL LETTER N WITH CIRCUMFLEX BELOW
4505	<o?'>	<U1E4C>	LATIN CAPITAL LETTER O WITH TILDE AND ACUTE
4506	<o?'>	<U1E4D>	LATIN SMALL LETTER O WITH TILDE AND ACUTE
4507	<o?:>	<U1E4E>	LATIN CAPITAL LETTER O WITH TILDE AND DIAERESIS
4508	<o?:>	<U1E4F>	LATIN SMALL LETTER O WITH TILDE AND DIAERESIS
4509	<o-'>	<U1E50>	LATIN CAPITAL LETTER O WITH MACRON AND GRAVE
4510	<o-'>	<U1E51>	LATIN SMALL LETTER O WITH MACRON AND GRAVE
4511	<o-'>	<U1E52>	LATIN CAPITAL LETTER O WITH MACRON AND ACUTE
4512	<o-'>	<U1E53>	LATIN SMALL LETTER O WITH MACRON AND ACUTE
4513	<p:'>	<U1E54>	LATIN CAPITAL LETTER P WITH ACUTE
4514	<p:'>	<U1E55>	LATIN SMALL LETTER P WITH ACUTE
4515	<p:>	<U1E56>	LATIN CAPITAL LETTER P WITH DOT ABOVE
4516	<p:>	<U1E57>	LATIN SMALL LETTER P WITH DOT ABOVE
4517	<r:>	<U1E58>	LATIN CAPITAL LETTER R WITH DOT ABOVE
4518	<r:>	<U1E59>	LATIN SMALL LETTER R WITH DOT ABOVE
4519	<r-'>	<U1E5A>	LATIN CAPITAL LETTER R WITH DOT BELOW
4520	<r-'>	<U1E5B>	LATIN SMALL LETTER R WITH DOT BELOW
4521	<r-->	<U1E5C>	LATIN CAPITAL LETTER R WITH DOT BELOW AND MACRON
4522	<r-->	<U1E5D>	LATIN SMALL LETTER R WITH DOT BELOW AND MACRON
4523	<r_>	<U1E5E>	LATIN CAPITAL LETTER R WITH LINE BELOW
4524	<r_>	<U1E5F>	LATIN SMALL LETTER R WITH LINE BELOW
4525	<s:>	<U1E60>	LATIN CAPITAL LETTER S WITH DOT ABOVE
4526	<s:>	<U1E61>	LATIN SMALL LETTER S WITH DOT ABOVE
4527	<s-'>	<U1E62>	LATIN CAPITAL LETTER S WITH DOT BELOW
4528	<s-'>	<U1E63>	LATIN SMALL LETTER S WITH DOT BELOW
4529	<s'>	<U1E64>	LATIN CAPITAL LETTER S WITH ACUTE AND DOT ABOVE
4530	<s'>	<U1E65>	LATIN SMALL LETTER S WITH ACUTE AND DOT ABOVE
4531	<s<,>	<U1E66>	LATIN CAPITAL LETTER S WITH CARON AND DOT ABOVE
4532	<s<,>	<U1E67>	LATIN SMALL LETTER S WITH CARON AND DOT ABOVE
4533	<s-.>	<U1E68>	LATIN CAPITAL LETTER S WITH DOT BELOW AND DOT ABOVE
4534	<s-.>	<U1E69>	LATIN SMALL LETTER S WITH DOT BELOW AND DOT ABOVE
4535	<t:>	<U1E6A>	LATIN CAPITAL LETTER T WITH DOT ABOVE
4536	<t:>	<U1E6B>	LATIN SMALL LETTER T WITH DOT ABOVE
4537	<t-'>	<U1E6C>	LATIN CAPITAL LETTER T WITH DOT BELOW
4538	<t-'>	<U1E6D>	LATIN SMALL LETTER T WITH DOT BELOW
4539	<t_>	<U1E6E>	LATIN CAPITAL LETTER T WITH LINE BELOW
4540	<t_>	<U1E6F>	LATIN SMALL LETTER T WITH LINE BELOW
4541	<t-/>	<U1E70>	LATIN CAPITAL LETTER T WITH CIRCUMFLEX BELOW
4542	<t-/>	<U1E71>	LATIN SMALL LETTER T WITH CIRCUMFLEX BELOW
4543	<u-->	<U1E72>	LATIN CAPITAL LETTER U WITH DIAERESIS BELOW
4544	<u-->	<U1E73>	LATIN SMALL LETTER U WITH DIAERESIS BELOW
4545	<u-?>	<U1E74>	LATIN CAPITAL LETTER U WITH TILDE BELOW
4546	<u-?>	<U1E75>	LATIN SMALL LETTER U WITH TILDE BELOW
4547	<u-/>	<U1E76>	LATIN CAPITAL LETTER U WITH CIRCUMFLEX BELOW
4548	<u-/>	<U1E77>	LATIN SMALL LETTER U WITH CIRCUMFLEX BELOW
4549	<u?'>	<U1E78>	LATIN CAPITAL LETTER U WITH TILDE AND ACUTE
4550	<u?'>	<U1E79>	LATIN SMALL LETTER U WITH TILDE AND ACUTE
4551	<u-: >	<U1E7A>	LATIN CAPITAL LETTER U WITH MACRON AND DIAERESIS
4552	<u-: >	<U1E7B>	LATIN SMALL LETTER U WITH MACRON AND DIAERESIS
4553	<v?'>	<U1E7C>	LATIN CAPITAL LETTER V WITH TILDE
4554	<v?'>	<U1E7D>	LATIN SMALL LETTER V WITH TILDE
4555	<v-: >	<U1E7E>	LATIN CAPITAL LETTER V WITH DOT BELOW
4556	<v-: >	<U1E7F>	LATIN SMALL LETTER V WITH DOT BELOW
4557	<w!'>	<U1E80>	LATIN CAPITAL LETTER W WITH GRAVE
4558	<w!'>	<U1E81>	LATIN SMALL LETTER W WITH GRAVE
4559	<w'>	<U1E82>	LATIN CAPITAL LETTER W WITH ACUTE
4560	<w'>	<U1E83>	LATIN SMALL LETTER W WITH ACUTE
4561	<w:>	<U1E84>	LATIN CAPITAL LETTER W WITH DIAERESIS
4562	<w:>	<U1E85>	LATIN SMALL LETTER W WITH DIAERESIS

4563	<W.>	<U1E86>	LATIN CAPITAL LETTER W WITH DOT ABOVE
4564	<w.>	<U1E87>	LATIN SMALL LETTER W WITH DOT ABOVE
4565	<W-.>	<U1E88>	LATIN CAPITAL LETTER W WITH DOT BELOW
4566	<w-.>	<U1E89>	LATIN SMALL LETTER W WITH DOT BELOW
4567	<X.>	<U1E8A>	LATIN CAPITAL LETTER X WITH DOT ABOVE
4568	<x.>	<U1E8B>	LATIN SMALL LETTER X WITH DOT ABOVE
4569	<X:>	<U1E8C>	LATIN CAPITAL LETTER X WITH DIAERESIS
4570	<x:>	<U1E8D>	LATIN SMALL LETTER X WITH DIAERESIS
4571	<Y.>	<U1E8E>	LATIN CAPITAL LETTER Y WITH DOT ABOVE
4572	<y.>	<U1E8F>	LATIN SMALL LETTER Y WITH DOT ABOVE
4573	<Z/>	<U1E90>	LATIN CAPITAL LETTER Z WITH CIRCUMFLEX
4574	<z/>	<U1E91>	LATIN SMALL LETTER Z WITH CIRCUMFLEX
4575	<Z-.>	<U1E92>	LATIN CAPITAL LETTER Z WITH DOT BELOW
4576	<z-.>	<U1E93>	LATIN SMALL LETTER Z WITH DOT BELOW
4577	<Z_>	<U1E94>	LATIN CAPITAL LETTER Z WITH LINE BELOW
4578	<z_>	<U1E95>	LATIN SMALL LETTER Z WITH LINE BELOW
4579	<A-.>	<U1EA0>	LATIN CAPITAL LETTER A WITH DOT BELOW
4580	<a-.>	<U1EA1>	LATIN SMALL LETTER A WITH DOT BELOW
4581	<A2>	<U1EA2>	LATIN CAPITAL LETTER A WITH HOOK ABOVE
4582	<a2>	<U1EA3>	LATIN SMALL LETTER A WITH HOOK ABOVE
4583	<A/>'>	<U1EA4>	LATIN CAPITAL LETTER A WITH CIRCUMFLEX AND ACUTE
4584	<a/>'>	<U1EA5>	LATIN SMALL LETTER A WITH CIRCUMFLEX AND ACUTE
4585	<A/>!>	<U1EA6>	LATIN CAPITAL LETTER A WITH CIRCUMFLEX AND GRAVE
4586	<a/>!>	<U1EA7>	LATIN SMALL LETTER A WITH CIRCUMFLEX AND GRAVE
4587	<A/>2>	<U1EA8>	LATIN CAPITAL LETTER A WITH CIRCUMFLEX AND HOOK ABOVE
4588	<a/>2>	<U1EA9>	LATIN SMALL LETTER A WITH CIRCUMFLEX AND HOOK ABOVE
4589	<A/>?>	<U1EAA>	LATIN CAPITAL LETTER A WITH CIRCUMFLEX AND TILDE
4590	<a/>?>	<U1EAB>	LATIN SMALL LETTER A WITH CIRCUMFLEX AND TILDE
4591	<A/>->	<U1EAC>	LATIN CAPITAL LETTER A WITH CIRCUMFLEX AND DOT BELOW
4592	<a/>->	<U1EAD>	LATIN SMALL LETTER A WITH CIRCUMFLEX AND DOT BELOW
4593	<A(')>	<U1EAE>	LATIN CAPITAL LETTER A WITH BREVE AND ACUTE
4594	<a(')>	<U1EAF>	LATIN SMALL LETTER A WITH BREVE AND ACUTE
4595	<A(!)>	<U1EB0>	LATIN CAPITAL LETTER A WITH BREVE AND GRAVE
4596	<a(!)>	<U1EB1>	LATIN SMALL LETTER A WITH BREVE AND GRAVE
4597	<A(2)>	<U1EB2>	LATIN CAPITAL LETTER A WITH BREVE AND HOOK ABOVE
4598	<a(2)>	<U1EB3>	LATIN SMALL LETTER A WITH BREVE AND HOOK ABOVE
4599	<A(?)>	<U1EB4>	LATIN CAPITAL LETTER A WITH BREVE AND TILDE
4600	<a(?)>	<U1EB5>	LATIN SMALL LETTER A WITH BREVE AND TILDE
4601	<A(-)>	<U1EB6>	LATIN CAPITAL LETTER A WITH BREVE AND DOT BELOW
4602	<a(-)>	<U1EB7>	LATIN SMALL LETTER A WITH BREVE AND DOT BELOW
4603	<E->	<U1EB8>	LATIN CAPITAL LETTER E WITH DOT BELOW
4604	<e->	<U1EB9>	LATIN SMALL LETTER E WITH DOT BELOW
4605	<E2>	<U1EBA>	LATIN CAPITAL LETTER E WITH HOOK ABOVE
4606	<e2>	<U1EBB>	LATIN SMALL LETTER E WITH HOOK ABOVE
4607	<E?>	<U1EBC>	LATIN CAPITAL LETTER E WITH TILDE
4608	<e?>	<U1EBD>	LATIN SMALL LETTER E WITH TILDE
4609	<E/>'>	<U1EBE>	LATIN CAPITAL LETTER E WITH CIRCUMFLEX AND ACUTE
4610	<e/>'>	<U1EBF>	LATIN SMALL LETTER E WITH CIRCUMFLEX AND ACUTE
4611	<E/>!>	<U1EC0>	LATIN CAPITAL LETTER E WITH CIRCUMFLEX AND GRAVE
4612	<e/>!>	<U1EC1>	LATIN SMALL LETTER E WITH CIRCUMFLEX AND GRAVE
4613	<E/>2>	<U1EC2>	LATIN CAPITAL LETTER E WITH CIRCUMFLEX AND HOOK ABOVE
4614	<e/>2>	<U1EC3>	LATIN SMALL LETTER E WITH CIRCUMFLEX AND HOOK ABOVE
4615	<E/>?>	<U1EC4>	LATIN CAPITAL LETTER E WITH CIRCUMFLEX AND TILDE
4616	<e/>?>	<U1EC5>	LATIN SMALL LETTER E WITH CIRCUMFLEX AND TILDE
4617	<E/>->	<U1EC6>	LATIN CAPITAL LETTER E WITH CIRCUMFLEX AND DOT BELOW
4618	<e/>->	<U1EC7>	LATIN SMALL LETTER E WITH CIRCUMFLEX AND DOT BELOW
4619	<i2>	<U1EC8>	LATIN CAPITAL LETTER I WITH HOOK ABOVE
4620	<i2>	<U1EC9>	LATIN SMALL LETTER I WITH HOOK ABOVE
4621	<I->	<U1ECA>	LATIN CAPITAL LETTER I WITH DOT BELOW
4622	<i->	<U1ECB>	LATIN SMALL LETTER I WITH DOT BELOW
4623	<o->	<U1ECC>	LATIN CAPITAL LETTER O WITH DOT BELOW
4624	<o->	<U1ECD>	LATIN SMALL LETTER O WITH DOT BELOW
4625	<o2>	<U1ECE>	LATIN CAPITAL LETTER O WITH HOOK ABOVE
4626	<o2>	<U1ECF>	LATIN SMALL LETTER O WITH HOOK ABOVE
4627	<O/>'>	<U1ED0>	LATIN CAPITAL LETTER O WITH CIRCUMFLEX AND ACUTE
4628	<o/>'>	<U1ED1>	LATIN SMALL LETTER O WITH CIRCUMFLEX AND ACUTE
4629	<O/>1>	<U1ED2>	LATIN CAPITAL LETTER O WITH CIRCUMFLEX AND GRAVE
4630	<o/>1>	<U1ED3>	LATIN SMALL LETTER O WITH CIRCUMFLEX AND GRAVE
4631	<O/>2>	<U1ED4>	LATIN CAPITAL LETTER O WITH CIRCUMFLEX AND HOOK ABOVE
4632	<o/>2>	<U1ED5>	LATIN SMALL LETTER O WITH CIRCUMFLEX AND HOOK ABOVE
4633	<O/>?>	<U1ED6>	LATIN CAPITAL LETTER O WITH CIRCUMFLEX AND TILDE
4634	<o/>?>	<U1ED7>	LATIN SMALL LETTER O WITH CIRCUMFLEX AND TILDE
4635	<O/>->	<U1ED8>	LATIN CAPITAL LETTER O WITH CIRCUMFLEX AND DOT BELOW
4636	<o/>->	<U1ED9>	LATIN SMALL LETTER O WITH CIRCUMFLEX AND DOT BELOW
4637	<O9'>	<U1EDA>	LATIN CAPITAL LETTER O WITH HORN AND ACUTE
4638	<O9'>	<U1EDB>	LATIN SMALL LETTER O WITH HORN AND ACUTE
4639	<O9!>	<U1EDC>	LATIN CAPITAL LETTER O WITH HORN AND GRAVE
4640	<o9!>	<U1EDD>	LATIN SMALL LETTER O WITH HORN AND GRAVE
4641	<O92>	<U1EDE>	LATIN CAPITAL LETTER O WITH HORN AND HOOK ABOVE
4642	<o92>	<U1EDF>	LATIN SMALL LETTER O WITH HORN AND HOOK ABOVE
4643	<O9?>	<U1EE0>	LATIN CAPITAL LETTER O WITH HORN AND TILDE
4644	<o9?>	<U1EE1>	LATIN SMALL LETTER O WITH HORN AND TILDE
4645	<O9->	<U1EE2>	LATIN CAPITAL LETTER O WITH HORN AND DOT BELOW
4646	<o9->	<U1EE3>	LATIN SMALL LETTER O WITH HORN AND DOT BELOW
4647	<U->	<U1EE4>	LATIN CAPITAL LETTER U WITH DOT BELOW
4648	<u->	<U1EE5>	LATIN SMALL LETTER U WITH DOT BELOW
4649	<U2>	<U1EE6>	LATIN CAPITAL LETTER U WITH HOOK ABOVE
4650	<u2>	<U1EE7>	LATIN SMALL LETTER U WITH HOOK ABOVE

4651	<U9'>	<U1EE8>	LATIN CAPITAL LETTER U WITH HORN AND ACUTE
4652	<U9'>	<U1EE9>	LATIN SMALL LETTER U WITH HORN AND ACUTE
4653	<U9!>	<U1EEA>	LATIN CAPITAL LETTER U WITH HORN AND GRAVE
4654	<u9!>	<U1EEB>	LATIN SMALL LETTER U WITH HORN AND GRAVE
4655	<U92>	<U1EEC>	LATIN CAPITAL LETTER U WITH HORN AND HOOK ABOVE
4656	<u92>	<U1EED>	LATIN SMALL LETTER U WITH HORN AND HOOK ABOVE
4657	<U9?>	<U1EEE>	LATIN CAPITAL LETTER U WITH HORN AND TILDE
4658	<u9?>	<U1EEF>	LATIN SMALL LETTER U WITH HORN AND TILDE
4659	<U9-.>	<U1EF0>	LATIN CAPITAL LETTER U WITH HORN AND DOT BELOW
4660	<u9-.>	<U1EF1>	LATIN SMALL LETTER U WITH HORN AND DOT BELOW
4661	<Y!>	<U1EF2>	LATIN CAPITAL LETTER Y WITH GRAVE
4662	<y!>	<U1EF3>	LATIN SMALL LETTER Y WITH GRAVE
4663	<Y-.>	<U1EF4>	LATIN CAPITAL LETTER Y WITH DOT BELOW
4664	<y-.>	<U1EF5>	LATIN SMALL LETTER Y WITH DOT BELOW
4665	<Y2>	<U1EF6>	LATIN CAPITAL LETTER Y WITH HOOK ABOVE
4666	<y2>	<U1EF7>	LATIN SMALL LETTER Y WITH HOOK ABOVE
4667	<Y?>	<U1EF8>	LATIN CAPITAL LETTER Y WITH TILDE
4668	<y?>	<U1EF9>	LATIN SMALL LETTER Y WITH TILDE
4669	<a*,>	<U1F00>	GREEK SMALL LETTER ALPHA WITH PSILI
4670	<a*;*>	<U1F01>	GREEK SMALL LETTER ALPHA WITH DASIA
4671	<a*,!>	<U1F02>	GREEK SMALL LETTER ALPHA WITH PSILI AND VARIA
4672	<a*;!>	<U1F03>	GREEK SMALL LETTER ALPHA WITH DASIA AND VARIA
4673	<a*,'/>	<U1F04>	GREEK SMALL LETTER ALPHA WITH PSILI AND OXIA
4674	<a*,'/>	<U1F05>	GREEK SMALL LETTER ALPHA WITH DASIA AND OXIA
4675	<a*,?>	<U1F06>	GREEK SMALL LETTER ALPHA WITH PSILI AND PERISPOMENI
4676	<a*;?>	<U1F07>	GREEK SMALL LETTER ALPHA WITH DASIA AND PERISPOMENI
4677	<A*,>	<U1F08>	GREEK CAPITAL LETTER ALPHA WITH PSILI
4678	<A*;*>	<U1F09>	GREEK CAPITAL LETTER ALPHA WITH DASIA
4679	<A*,!>	<U1F0A>	GREEK CAPITAL LETTER ALPHA WITH PSILI AND VARIA
4680	<A*;!>	<U1F0B>	GREEK CAPITAL LETTER ALPHA WITH DASIA AND VARIA
4681	<A*,'/>	<U1F0C>	GREEK CAPITAL LETTER ALPHA WITH PSILI AND OXIA
4682	<A*,'/>	<U1F0D>	GREEK CAPITAL LETTER ALPHA WITH DASIA AND OXIA
4683	<A*,?>	<U1F0E>	GREEK CAPITAL LETTER ALPHA WITH PSILI AND PERISPOMENI
4684	<A*;?>	<U1F0F>	GREEK CAPITAL LETTER ALPHA WITH DASIA AND PERISPOMENI
4685	<e*,>	<U1F10>	GREEK SMALL LETTER EPSILON WITH PSILI
4686	<e*;*>	<U1F11>	GREEK SMALL LETTER EPSILON WITH DASIA
4687	<e*,!>	<U1F12>	GREEK SMALL LETTER EPSILON WITH PSILI AND VARIA
4688	<e*;!>	<U1F13>	GREEK SMALL LETTER EPSILON WITH DASIA AND VARIA
4689	<e*,'/>	<U1F14>	GREEK SMALL LETTER EPSILON WITH PSILI AND OXIA
4690	<e*,'/>	<U1F15>	GREEK SMALL LETTER EPSILON WITH DASIA AND OXIA
4691	<E*,>	<U1F18>	GREEK CAPITAL LETTER EPSILON WITH PSILI
4692	<E*;*>	<U1F19>	GREEK CAPITAL LETTER EPSILON WITH DASIA
4693	<E*,!>	<U1F1A>	GREEK CAPITAL LETTER EPSILON WITH PSILI AND VARIA
4694	<E*;!>	<U1F1B>	GREEK CAPITAL LETTER EPSILON WITH DASIA AND VARIA
4695	<E*,'/>	<U1F1C>	GREEK CAPITAL LETTER EPSILON WITH PSILI AND OXIA
4696	<E*,'/>	<U1F1D>	GREEK CAPITAL LETTER EPSILON WITH DASIA AND OXIA
4697	<Y*,>	<U1F20>	GREEK SMALL LETTER ETA WITH PSILI
4698	<y*;*>	<U1F21>	GREEK SMALL LETTER ETA WITH DASIA
4699	<y*;!>	<U1F22>	GREEK SMALL LETTER ETA WITH PSILI AND VARIA
4700	<y*;!>	<U1F23>	GREEK SMALL LETTER ETA WITH DASIA AND VARIA
4701	<y*,'/>	<U1F24>	GREEK SMALL LETTER ETA WITH PSILI AND OXIA
4702	<y*,'/>	<U1F25>	GREEK SMALL LETTER ETA WITH DASIA AND OXIA
4703	<y*,?>	<U1F26>	GREEK SMALL LETTER ETA WITH PSILI AND PERISPOMENI
4704	<y*;?>	<U1F27>	GREEK SMALL LETTER ETA WITH DASIA AND PERISPOMENI
4705	<Y*,>	<U1F28>	GREEK CAPITAL LETTER ETA WITH PSILI
4706	<Y*;*>	<U1F29>	GREEK CAPITAL LETTER ETA WITH DASIA
4707	<Y*,!>	<U1F2A>	GREEK CAPITAL LETTER ETA WITH PSILI AND VARIA
4708	<Y*;!>	<U1F2B>	GREEK CAPITAL LETTER ETA WITH DASIA AND VARIA
4709	<Y*,'/>	<U1F2C>	GREEK CAPITAL LETTER ETA WITH PSILI AND OXIA
4710	<Y*,'/>	<U1F2D>	GREEK CAPITAL LETTER ETA WITH DASIA AND OXIA
4711	<Y*,?>	<U1F2E>	GREEK CAPITAL LETTER ETA WITH PSILI AND PERISPOMENI
4712	<Y*;?>	<U1F2F>	GREEK CAPITAL LETTER ETA WITH DASIA AND PERISPOMENI
4713	<i*,>	<U1F30>	GREEK SMALL LETTER IOTA WITH PSILI
4714	<i*;*>	<U1F31>	GREEK SMALL LETTER IOTA WITH DASIA
4715	<i*,!>	<U1F32>	GREEK SMALL LETTER IOTA WITH PSILI AND VARIA
4716	<i*;!>	<U1F33>	GREEK SMALL LETTER IOTA WITH DASIA AND VARIA
4717	<i*,'/>	<U1F34>	GREEK SMALL LETTER IOTA WITH PSILI AND OXIA
4718	<i*,'/>	<U1F35>	GREEK SMALL LETTER IOTA WITH DASIA AND OXIA
4719	<i*,?>	<U1F36>	GREEK SMALL LETTER IOTA WITH PSILI AND PERISPOMENI
4720	<i*;?>	<U1F37>	GREEK SMALL LETTER IOTA WITH DASIA AND PERISPOMENI
4721	<I*,>	<U1F38>	GREEK CAPITAL LETTER IOTA WITH PSILI
4722	<I*;*>	<U1F39>	GREEK CAPITAL LETTER IOTA WITH DASIA
4723	<I*,!>	<U1F3A>	GREEK CAPITAL LETTER IOTA WITH PSILI AND VARIA
4724	<I*;!>	<U1F3B>	GREEK CAPITAL LETTER IOTA WITH DASIA AND VARIA
4725	<I*,'/>	<U1F3C>	GREEK CAPITAL LETTER IOTA WITH PSILI AND OXIA
4726	<I*,'/>	<U1F3D>	GREEK CAPITAL LETTER IOTA WITH DASIA AND OXIA
4727	<I*,?>	<U1F3E>	GREEK CAPITAL LETTER IOTA WITH PSILI AND PERISPOMENI
4728	<I*;?>	<U1F3F>	GREEK CAPITAL LETTER IOTA WITH DASIA AND PERISPOMENI
4729	<o*,>	<U1F40>	GREEK SMALL LETTER OMICRON WITH PSILI
4730	<o*;*>	<U1F41>	GREEK SMALL LETTER OMICRON WITH DASIA
4731	<o*,!>	<U1F42>	GREEK SMALL LETTER OMICRON WITH PSILI AND VARIA
4732	<o*;!>	<U1F43>	GREEK SMALL LETTER OMICRON WITH DASIA AND VARIA
4733	<o*,'/>	<U1F44>	GREEK SMALL LETTER OMICRON WITH PSILI AND OXIA
4734	<o*,'/>	<U1F45>	GREEK SMALL LETTER OMICRON WITH DASIA AND OXIA
4735	<o*,>	<U1F48>	GREEK CAPITAL LETTER OMICRON WITH PSILI
4736	<o*;*>	<U1F49>	GREEK CAPITAL LETTER OMICRON WITH DASIA
4737	<o*,!>	<U1F4A>	GREEK CAPITAL LETTER OMICRON WITH PSILI AND VARIA
4738	<o*;!>	<U1F4B>	GREEK CAPITAL LETTER OMICRON WITH DASIA AND VARIA

4739	<O*, ' >	<U1F4C>	GREEK CAPITAL LETTER OMEGON WITH PSILI AND OXIA
4740	<O*, ' >	<U1F4D>	GREEK CAPITAL LETTER OMEGON WITH DASIA AND OXIA
4741	<u*, ' >	<U1F50>	GREEK SMALL LETTER UPSILON WITH PSILI
4742	<u*, ' >	<U1F51>	GREEK SMALL LETTER UPSILON WITH DASIA
4743	<u*, ! >	<U1F52>	GREEK SMALL LETTER UPSILON WITH PSILI AND VARIA
4744	<u*, ! >	<U1F53>	GREEK SMALL LETTER UPSILON WITH DASIA AND VARIA
4745	<u*, ' >	<U1F54>	GREEK SMALL LETTER UPSILON WITH PSILI AND OXIA
4746	<u*, ' >	<U1F55>	GREEK SMALL LETTER UPSILON WITH DASIA AND OXIA
4747	<u*, ? >	<U1F56>	GREEK SMALL LETTER UPSILON WITH PSILI AND PERISPOMENI
4748	<u*, ? >	<U1F57>	GREEK SMALL LETTER UPSILON WITH DASIA AND PERISPOMENI
4749	<U*, ' >	<U1F59>	GREEK CAPITAL LETTER UPSILON WITH DASIA
4750	<U*, ! >	<U1F5B>	GREEK CAPITAL LETTER UPSILON WITH DASIA AND VARIA
4751	<U*, ' >	<U1F5D>	GREEK CAPITAL LETTER UPSILON WITH DASIA AND OXIA
4752	<U*, ? >	<U1F5F>	GREEK CAPITAL LETTER UPSILON WITH DASIA AND PERISPOMENI
4753	<w*, ' >	<U1F60>	GREEK SMALL LETTER OMEGA WITH PSILI
4754	<w*, ' >	<U1F61>	GREEK SMALL LETTER OMEGA WITH DASIA
4755	<w*, ! >	<U1F62>	GREEK SMALL LETTER OMEGA WITH PSILI AND VARIA
4756	<w*, ! >	<U1F63>	GREEK SMALL LETTER OMEGA WITH DASIA AND VARIA
4757	<w*, ' >	<U1F64>	GREEK SMALL LETTER OMEGA WITH PSILI AND OXIA
4758	<w*, ' >	<U1F65>	GREEK SMALL LETTER OMEGA WITH DASIA AND OXIA
4759	<w*, ? >	<U1F66>	GREEK SMALL LETTER OMEGA WITH PSILI AND PERISPOMENI
4760	<w*, ? >	<U1F67>	GREEK SMALL LETTER OMEGA WITH DASIA AND PERISPOMENI
4761	<w*, ' >	<U1F68>	GREEK CAPITAL LETTER OMEGA WITH PSILI
4762	<w*, ' >	<U1F69>	GREEK CAPITAL LETTER OMEGA WITH DASIA
4763	<w*, ! >	<U1F6A>	GREEK CAPITAL LETTER OMEGA WITH PSILI AND VARIA
4764	<w*, ! >	<U1F6B>	GREEK CAPITAL LETTER OMEGA WITH DASIA AND VARIA
4765	<w*, ' >	<U1F6C>	GREEK CAPITAL LETTER OMEGA WITH PSILI AND OXIA
4766	<w*, ' >	<U1F6D>	GREEK CAPITAL LETTER OMEGA WITH DASIA AND OXIA
4767	<w*, ? >	<U1F6E>	GREEK CAPITAL LETTER OMEGA WITH PSILI AND PERISPOMENI
4768	<w*, ? >	<U1F6F>	GREEK CAPITAL LETTER OMEGA WITH DASIA AND PERISPOMENI
4769	<a*! >	<U1F70>	GREEK SMALL LETTER ALPHA WITH VARIA
4770	<a*' >	<U1F71>	GREEK SMALL LETTER ALPHA WITH OXIA
4771	<e*! >	<U1F72>	GREEK SMALL LETTER EPSILON WITH VARIA
4772	<e*' >	<U1F73>	GREEK SMALL LETTER EPSILON WITH OXIA
4773	<y*! >	<U1F74>	GREEK SMALL LETTER ETA WITH VARIA
4774	<y*' >	<U1F75>	GREEK SMALL LETTER ETA WITH OXIA
4775	<i*! >	<U1F76>	GREEK SMALL LETTER IOTA WITH VARIA
4776	<i*! >	<U1F77>	GREEK SMALL LETTER IOTA WITH OXIA
4777	<o*! >	<U1F78>	GREEK SMALL LETTER OMICRON WITH VARIA
4778	<o*' >	<U1F79>	GREEK SMALL LETTER OMICRON WITH OXIA
4779	<u*! >	<U1F7A>	GREEK SMALL LETTER UPSILON WITH VARIA
4780	<u*' >	<U1F7B>	GREEK SMALL LETTER UPSILON WITH OXIA
4781	<w*! >	<U1F7C>	GREEK SMALL LETTER OMEGA WITH VARIA
4782	<w*' >	<U1F7D>	GREEK SMALL LETTER OMEGA WITH OXIA
4783	<a*, j >	<U1F80>	GREEK SMALL LETTER ALPHA WITH PSILI AND YPOGEGRAMMENI
4784	<a*; j >	<U1F81>	GREEK SMALL LETTER ALPHA WITH DASIA AND YPOGEGRAMMENI
4785	<a*, ! j >	<U1F82>	GREEK SMALL LETTER ALPHA WITH PSILI AND VARIA AND YPOGEGRAMMENI
4786	<a*; ! j >	<U1F83>	GREEK SMALL LETTER ALPHA WITH DASIA AND VARIA AND YPOGEGRAMMENI
4787	<a*, ' j >	<U1F84>	GREEK SMALL LETTER ALPHA WITH PSILI AND OXIA AND YPOGEGRAMMENI
4788	<a*, ' j >	<U1F85>	GREEK SMALL LETTER ALPHA WITH DASIA AND OXIA AND YPOGEGRAMMENI
4789	<a*, ? j >	<U1F86>	GREEK SMALL LETTER ALPHA WITH PSILI AND PERISPOMENI AND YPOGEGRAMMENI
4790	<a*; ? j >	<U1F87>	GREEK SMALL LETTER ALPHA WITH DASIA AND PERISPOMENI AND YPOGEGRAMMENI
4791	<A*, J >	<U1F88>	GREEK CAPITAL LETTER ALPHA WITH PSILI AND PROSGEGRAMMENI
4792	<A*; J >	<U1F89>	GREEK CAPITAL LETTER ALPHA WITH DASIA AND PROSGEGRAMMENI
4793	<A*, ! J >	<U1F8A>	GREEK CAPITAL LETTER ALPHA WITH PSILI AND VARIA AND PROSGEGRAMMENI
4794	<A*, ! J >	<U1F8B>	GREEK CAPITAL LETTER ALPHA WITH DASIA AND VARIA AND PROSGEGRAMMENI
4795	<A*, ' J >	<U1F8C>	GREEK CAPITAL LETTER ALPHA WITH PSILI AND OXIA AND PROSGEGRAMMENI
4796	<A*, ' J >	<U1F8D>	GREEK CAPITAL LETTER ALPHA WITH DASIA AND OXIA AND PROSGEGRAMMENI
4797	<A*, ? J >	<U1F8E>	GREEK CAPITAL LETTER ALPHA WITH PSILI AND PERISPOMENI AND PROSGEGRAMMENI
4798	<A*; ? J >	<U1F8F>	GREEK CAPITAL LETTER ALPHA WITH DASIA AND PERISPOMENI AND PROSGEGRAMMENI
4799	<y*, ' j >	<U1F90>	GREEK SMALL LETTER ETA WITH PSILI AND YPOGEGRAMMENI
4800	<y*, ' j >	<U1F91>	GREEK SMALL LETTER ETA WITH DASIA AND YPOGEGRAMMENI
4801	<y*, ! j >	<U1F92>	GREEK SMALL LETTER ETA WITH PSILI AND VARIA AND YPOGEGRAMMENI
4802	<y*, ! j >	<U1F93>	GREEK SMALL LETTER ETA WITH DASIA AND VARIA AND YPOGEGRAMMENI
4803	<y*, ' j >	<U1F94>	GREEK SMALL LETTER ETA WITH PSILI AND OXIA AND YPOGEGRAMMENI
4804	<y*, ' j >	<U1F95>	GREEK SMALL LETTER ETA WITH DASIA AND OXIA AND YPOGEGRAMMENI
4805	<y*, ? j >	<U1F96>	GREEK SMALL LETTER ETA WITH PSILI AND PERISPOMENI AND YPOGEGRAMMENI
4806	<y*, ? j >	<U1F97>	GREEK SMALL LETTER ETA WITH DASIA AND PERISPOMENI AND YPOGEGRAMMENI
4807	<Y*, J >	<U1F98>	GREEK CAPITAL LETTER ETA WITH PSILI AND PROSGEGRAMMENI
4808	<Y*, J >	<U1F99>	GREEK CAPITAL LETTER ETA WITH DASIA AND PROSGEGRAMMENI
4809	<Y*, ! J >	<U1F9A>	GREEK CAPITAL LETTER ETA WITH PSILI AND VARIA AND PROSGEGRAMMENI
4810	<Y*, ! J >	<U1F9B>	GREEK CAPITAL LETTER ETA WITH DASIA AND VARIA AND PROSGEGRAMMENI
4811	<Y*, ' J >	<U1F9C>	GREEK CAPITAL LETTER ETA WITH PSILI AND OXIA AND PROSGEGRAMMENI
4812	<Y*, ' J >	<U1F9D>	GREEK CAPITAL LETTER ETA WITH DASIA AND OXIA AND PROSGEGRAMMENI
4813	<Y*, ? J >	<U1F9E>	GREEK CAPITAL LETTER ETA WITH PSILI AND PERISPOMENI AND PROSGEGRAMMENI
4814	<Y*; ? J >	<U1F9F>	GREEK CAPITAL LETTER ETA WITH DASIA AND PERISPOMENI AND PROSGEGRAMMENI
4815	<w*, ' j >	<U1FA0>	GREEK SMALL LETTER OMEGA WITH PSILI AND YPOGEGRAMMENI
4816	<w*, ' j >	<U1FA1>	GREEK SMALL LETTER OMEGA WITH DASIA AND YPOGEGRAMMENI
4817	<w*, ! j >	<U1FA2>	GREEK SMALL LETTER OMEGA WITH PSILI AND VARIA AND YPOGEGRAMMENI
4818	<w*, ! j >	<U1FA3>	GREEK SMALL LETTER OMEGA WITH DASIA AND VARIA AND YPOGEGRAMMENI
4819	<w*, ' j >	<U1FA4>	GREEK SMALL LETTER OMEGA WITH PSILI AND OXIA AND YPOGEGRAMMENI
4820	<w*, ' j >	<U1FA5>	GREEK SMALL LETTER OMEGA WITH DASIA AND OXIA AND YPOGEGRAMMENI
4821	<w*, ? j >	<U1FA6>	GREEK SMALL LETTER OMEGA WITH PSILI AND PERISPOMENI AND YPOGEGRAMMENI
4822	<w*, ? j >	<U1FA7>	GREEK SMALL LETTER OMEGA WITH DASIA AND PERISPOMENI AND YPOGEGRAMMENI
4823	<w*, J >	<U1FA8>	GREEK CAPITAL LETTER OMEGA WITH PSILI AND PROSGEGRAMMENI
4824	<w*, J >	<U1FA9>	GREEK CAPITAL LETTER OMEGA WITH DASIA AND PROSGEGRAMMENI
4825	<w*, ! J >	<U1FAA>	GREEK CAPITAL LETTER OMEGA WITH PSILI AND VARIA AND PROSGEGRAMMENI
4826	<w*, ! J >	<U1FAB>	GREEK CAPITAL LETTER OMEGA WITH DASIA AND VARIA AND PROSGEGRAMMENI

4827	<W*, 'J>	<U1FAC>	GREEK CAPITAL LETTER OMEGA WITH PSILI AND OXIA AND PROSGEGRAMMENI
4828	<W*, 'J>	<U1FAD>	GREEK CAPITAL LETTER OMEGA WITH DASIA AND OXIA AND PROSGEGRAMMENI
4829	<W*, ?J>	<U1FAE>	GREEK CAPITAL LETTER OMEGA WITH PSILI AND PERISPOMENI AND PROSGEGRAMMENI
4830	<W*, ?J>	<U1FAF>	GREEK CAPITAL LETTER OMEGA WITH DASIA AND PERISPOMENI AND PROSGEGRAMMENI
4831	<a*(>	<U1FB0>	GREEK SMALL LETTER ALPHA WITH VRACHY
4832	<a*->	<U1FB1>	GREEK SMALL LETTER ALPHA WITH MACRON
4833	<a*!j>	<U1FB2>	GREEK SMALL LETTER ALPHA WITH VARIA AND YPOGEGRAMMENI
4834	<a*j>	<U1FB3>	GREEK SMALL LETTER ALPHA WITH YPOGEGRAMMENI
4835	<a*'j>	<U1FB4>	GREEK SMALL LETTER ALPHA WITH OXIA AND YPOGEGRAMMENI
4836	<a*?>	<U1FB6>	GREEK SMALL LETTER ALPHA WITH PERISPOMENI
4837	<a*?j>	<U1FB7>	GREEK SMALL LETTER ALPHA WITH PERISPOMENI AND YPOGEGRAMMENI
4838	<A*(>	<U1FB8>	GREEK CAPITAL LETTER ALPHA WITH VRACHY
4839	<A*->	<U1FB9>	GREEK CAPITAL LETTER ALPHA WITH MACRON
4840	<A*!>	<U1FBA>	GREEK CAPITAL LETTER ALPHA WITH VARIA
4841	<A*'>	<U1FBB>	GREEK CAPITAL LETTER ALPHA WITH OXIA
4842	<A*J>	<U1FBC>	GREEK CAPITAL LETTER ALPHA WITH PROSGEGRAMMENI
4843	<)*>	<U1FBD>	GREEK KORONIS
4844	<J3>	<U1FBE>	GREEK PROSGEGRAMMENI
4845	<, ,>	<U1FBF>	GREEK PSILI
4846	<?*>	<U1FC0>	GREEK PERISPOMENI
4847	<?:>	<U1FC1>	GREEK DIALYTIKA AND PERISPOMENI
4848	<y*!j>	<U1FC2>	GREEK SMALL LETTER ETA WITH VARIA AND YPOGEGRAMMENI
4849	<y*j>	<U1FC3>	GREEK SMALL LETTER ETA WITH YPOGEGRAMMENI
4850	<y*'j>	<U1FC4>	GREEK SMALL LETTER ETA WITH OXIA AND YPOGEGRAMMENI
4851	<y*?>	<U1FC6>	GREEK SMALL LETTER ETA WITH PERISPOMENI
4852	<y*?j>	<U1FC7>	GREEK SMALL LETTER ETA WITH PERISPOMENI AND YPOGEGRAMMENI
4853	<E*!!>	<U1FC8>	GREEK CAPITAL LETTER EPSILON WITH VARIA
4854	<E*'>	<U1FC9>	GREEK CAPITAL LETTER EPSILON WITH OXIA
4855	<Y*!>	<U1FCA>	GREEK CAPITAL LETTER ETA WITH VARIA
4856	<Y*'>	<U1FCB>	GREEK CAPITAL LETTER ETA WITH OXIA
4857	<Y*J>	<U1FCC>	GREEK CAPITAL LETTER ETA WITH PROSGEGRAMMENI
4858	<, !>	<U1FCD>	GREEK PSILI AND VARIA
4859	<, '>	<U1FCE>	GREEK PSILI AND OXIA
4860	<?, >	<U1FCF>	GREEK PSILI AND PERISPOMENI
4861	<i*(>	<U1FD0>	GREEK SMALL LETTER IOTA WITH VRACHY
4862	<i*->	<U1FD1>	GREEK SMALL LETTER IOTA WITH MACRON
4863	<i*!: !>	<U1FD2>	GREEK SMALL LETTER IOTA WITH DIALYTIKA AND VARIA
4864	<i*': '>	<U1FD3>	GREEK SMALL LETTER IOTA WITH DIALYTIKA AND OXIA
4865	<i*?>	<U1FD6>	GREEK SMALL LETTER IOTA WITH PERISPOMENI
4866	<i*?:>	<U1FD7>	GREEK SMALL LETTER IOTA WITH DIALYTIKA AND PERISPOMENI
4867	<I*(>	<U1FD8>	GREEK CAPITAL LETTER IOTA WITH VRACHY
4868	<I*->	<U1FD9>	GREEK CAPITAL LETTER IOTA WITH MACRON
4869	<I*!>	<U1FDA>	GREEK CAPITAL LETTER IOTA WITH VARIA
4870	<I*’>	<U1FDB>	GREEK CAPITAL LETTER IOTA WITH OXIA
4871	<; !>	<U1FDD>	GREEK DASIA AND VARIA
4872	<; '>	<U1FDE>	GREEK DASIA AND OXIA
4873	<?; >	<U1FDF>	GREEK DASIA AND PERISPOMENI
4874	<u*(>	<U1FE0>	GREEK SMALL LETTER UPSILON WITH VRACHY
4875	<u*->	<U1FE1>	GREEK SMALL LETTER UPSILON WITH MACRON
4876	<u*!: !>	<U1FE2>	GREEK SMALL LETTER UPSILON WITH DIALYTIKA AND VARIA
4877	<u*': '>	<U1FE3>	GREEK SMALL LETTER UPSILON WITH DIALYTIKA AND OXIA
4878	<r*,>	<U1FE4>	GREEK SMALL LETTER RHO WITH PSILI
4879	<r*; >	<U1FE5>	GREEK SMALL LETTER RHO WITH DASIA
4880	<u*?>	<U1FE6>	GREEK SMALL LETTER UPSILON WITH PERISPOMENI
4881	<u*?:>	<U1FE7>	GREEK SMALL LETTER UPSILON WITH DIALYTIKA AND PERISPOMENI
4882	<U*(>	<U1FE8>	GREEK CAPITAL LETTER UPSILON WITH VRACHY
4883	<U*->	<U1FE9>	GREEK CAPITAL LETTER UPSILON WITH MACRON
4884	<U*!>	<U1FEA>	GREEK CAPITAL LETTER UPSILON WITH VARIA
4885	<U*’>	<U1FEB>	GREEK CAPITAL LETTER UPSILON WITH OXIA
4886	<R*;/>	<U1FEC>	GREEK CAPITAL LETTER RHO WITH DASIA
4887	<!; :>	<U1FED>	GREEK DIALYTIKA AND VARIA
4888	<!: '>	<U1FEE>	GREEK DIALYTIKA AND OXIA
4889	<!;*>	<U1FEF>	GREEK VARIA
4890	<w*!j>	<U1FF2>	GREEK SMALL LETTER OMEGA WITH VARIA AND YPOGEGRAMMENI
4891	<w*j>	<U1FF3>	GREEK SMALL LETTER OMEGA WITH YPOGEGRAMMENI
4892	<w* j>	<U1FF4>	GREEK SMALL LETTER OMEGA WITH OXIA AND YPOGEGRAMMENI
4893	<w*?>	<U1FF6>	GREEK SMALL LETTER OMEGA WITH PERISPOMENI
4894	<w*?j>	<U1FF7>	GREEK SMALL LETTER OMEGA WITH PERISPOMENI AND YPOGEGRAMMENI
4895	<O*!>	<U1FF8>	GREEK CAPITAL LETTER OMICRON WITH VARIA
4896	<O*’>	<U1FF9>	GREEK CAPITAL LETTER OMICRON WITH OXIA
4897	<W*!>	<U1FFA>	GREEK CAPITAL LETTER OMEGA WITH VARIA
4898	<W*’>	<U1FFB>	GREEK CAPITAL LETTER OMEGA WITH OXIA
4899	<W*J>	<U1FFC>	GREEK CAPITAL LETTER OMEGA WITH PROSGEGRAMMENI
4900	<//,*>	<U1FFD>	GREEK OXIA
4901	<; ;>	<U1FFE>	GREEK DASIA
4902	<1N>	<U2002>	EN SPACE
4903	<1M>	<U2003>	EM SPACE
4904	<3M>	<U2004>	THREE-PER-EM SPACE
4905	<4M>	<U2005>	FOUR-PER-EM SPACE
4906	<6M>	<U2006>	SIX-PER-EM SPACE
4907	<LR>	<U200E>	LEFT-TO-RIGHT MARK
4908	<RL>	<U200F>	RIGHT-TO-LEFT MARK
4909	<1T>	<U2009>	THIN SPACE
4910	<1H>	<U200A>	HAIR SPACE
4911	<-1>	<U2010>	HYPHEN
4912	<-N>	<U2013>	EN DASH
4913	<-M>	<U2014>	EM DASH
4914	<-3>	<U2015>	HORIZONTAL BAR

4915	<!2>	<U2016>	DOUBLE VERTICAL LINE
4916	<=2>	<U2017>	DOUBLE LOW LINE
4917	<'6>	<U2018>	LEFT SINGLE QUOTATION MARK
4918	<'9>	<U2019>	RIGHT SINGLE QUOTATION MARK
4919	<.9>	<U201A>	SINGLE LOW-9 QUOTATION MARK
4920	<9'>	<U201B>	SINGLE HIGH-REVERSED-9 QUOTATION MARK
4921	<"6>	<U201C>	LEFT DOUBLE QUOTATION MARK
4922	<"9>	<U201D>	RIGHT DOUBLE QUOTATION MARK
4923	<:9>	<U201E>	DOUBLE LOW-9 QUOTATION MARK
4924	<9">	<U201F>	DOUBLE HIGH-REVERSED-9 QUOTATION MARK
4925	<//>	<U2020>	DAGGER
4926	<///>	<U2021>	DOUBLE DAGGER
4927	<sb>	<U2022>	BULLET
4928	<3b>	<U2023>	TRIANGULAR BULLET
4929	<..>	<U2025>	TWO DOT LEADER
4930	<.3>	<U2026>	HORIZONTAL ELLIPSIS
4931	<.->	<U2027>	HYPHENATION POINT
4932	<linesep>	<U2028>	LINE SEPARATOR
4933	<parsep>	<U2029>	PARAGRAPH SEPARATOR
4934	<%0>	<U2030>	PER MILLE SIGN
4935	<1'>	<U2032>	PRIME
4936	<2'>	<U2033>	DOUBLE PRIME
4937	<3'>	<U2034>	TRIPLE PRIME
4938	<1">	<U2035>	REVERSED PRIME
4939	<2">	<U2036>	REVERSED DOUBLE PRIME
4940	<3">	<U2037>	REVERSED TRIPLE PRIME
4941	<Ca>	<U2038>	CARET
4942	<<1>	<U2039>	SINGLE LEFT-POINTING ANGLE QUOTATION MARK
4943	</>1>	<U203A>	SINGLE RIGHT-POINTING ANGLE QUOTATION MARK
4944	<:X>	<U203B>	REFERENCE MARK
4945	<!*2>	<U203C>	DOUBLE EXCLAMATION MARK
4946	<'->	<U203E>	OVERLINE
4947	<-b>	<U2043>	HYPHEN BULLET
4948	</f>	<U2044>	FRACTION SLASH
4949	<0s>	<U2070>	SUPERSCRIPT ZERO
4950	<4s>	<U2074>	SUPERSCRIPT FOUR
4951	<5s>	<U2075>	SUPERSCRIPT FIVE
4952	<6s>	<U2076>	SUPERSCRIPT SIX
4953	<7s>	<U2077>	SUPERSCRIPT SEVEN
4954	<8s>	<U2078>	SUPERSCRIPT EIGHT
4955	<9s>	<U2079>	SUPERSCRIPT NINE
4956	<+s>	<U207A>	SUPERSCRIPT PLUS SIGN
4957	<-s>	<U207B>	SUPERSCRIPT MINUS
4958	<=s>	<U207C>	SUPERSCRIPT EQUALS SIGN
4959	<(s>	<U207D>	SUPERSCRIPT LEFT PARENTHESIS
4960	<)s>	<U207E>	SUPERSCRIPT RIGHT PARENTHESIS
4961	<nS>	<U207F>	SUPERSCRIPT LATIN SMALL LETTER N
4962	<0s>	<U2080>	SUBSCRIPT ZERO
4963	<1s>	<U2081>	SUBSCRIPT ONE
4964	<2s>	<U2082>	SUBSCRIPT TWO
4965	<3s>	<U2083>	SUBSCRIPT THREE
4966	<4s>	<U2084>	SUBSCRIPT FOUR
4967	<5s>	<U2085>	SUBSCRIPT FIVE
4968	<6s>	<U2086>	SUBSCRIPT SIX
4969	<7s>	<U2087>	SUBSCRIPT SEVEN
4970	<8s>	<U2088>	SUBSCRIPT EIGHT
4971	<9s>	<U2089>	SUBSCRIPT NINE
4972	<+s>	<U208A>	SUBSCRIPT PLUS SIGN
4973	<-s>	<U208B>	SUBSCRIPT MINUS
4974	<=s>	<U208C>	SUBSCRIPT EQUALS SIGN
4975	<(s>	<U208D>	SUBSCRIPT LEFT PARENTHESIS
4976	<)s>	<U208E>	SUBSCRIPT RIGHT PARENTHESIS
4977	<Ff>	<U20A3>	FRENCH FRANC SIGN
4978	<L1>	<U20A4>	LIRA SIGN
4979	<Pt>	<U20A7>	PESETA SIGN
4980	<W=>	<U20A9>	WON SIGN
4981	<"7>	<U20D1>	COMBINING RIGHT HARPOON ABOVE
4982	<oC>	<U2103>	DEGREE CELSIUS
4983	<co>	<U2105>	CARE OF
4984	<of>	<U2109>	DEGREE FAHRENHEIT
4985	<N0>	<U2116>	NUMERO SIGN
4986	<PO>	<U2117>	SOUND RECORDING COPYRIGHT
4987	<Rx>	<U211E>	PRESCRIPTION TAKE
4988	<SM>	<U2120>	SERVICE MARK
4989	<TM>	<U2122>	TRADE MARK SIGN
4990	<Om>	<U2126>	OHM SIGN
4991	<AO>	<U212B>	ANGSTROM SIGN
4992	<Est>	<U212E>	ESTIMATED SYMBOL
4993	<13>	<U2153>	VULGAR FRACTION ONE THIRD
4994	<23>	<U2154>	VULGAR FRACTION TWO THIRDS
4995	<15>	<U2155>	VULGAR FRACTION ONE FIFTH
4996	<25>	<U2156>	VULGAR FRACTION TWO FIFTHS
4997	<35>	<U2157>	VULGAR FRACTION THREE FIFTHS
4998	<45>	<U2158>	VULGAR FRACTION FOUR FIFTHS
4999	<16>	<U2159>	VULGAR FRACTION ONE SIXTH
5000	<56>	<U215A>	VULGAR FRACTION FIVE SIXTHS
5001	<18>	<U215B>	VULGAR FRACTION ONE EIGHTH
5002	<38>	<U215C>	VULGAR FRACTION THREE EIGHTHS

5003	<58>	<U215D>	VULGAR FRACTION FIVE EIGHTHS
5004	<78>	<U215E>	VULGAR FRACTION SEVEN EIGHTHS
5005	<1R>	<U2160>	ROMAN NUMERAL ONE
5006	<2R>	<U2161>	ROMAN NUMERAL TWO
5007	<3R>	<U2162>	ROMAN NUMERAL THREE
5008	<4R>	<U2163>	ROMAN NUMERAL FOUR
5009	<5R>	<U2164>	ROMAN NUMERAL FIVE
5010	<6R>	<U2165>	ROMAN NUMERAL SIX
5011	<7R>	<U2166>	ROMAN NUMERAL SEVEN
5012	<8R>	<U2167>	ROMAN NUMERAL EIGHT
5013	<9R>	<U2168>	ROMAN NUMERAL NINE
5014	<aR>	<U2169>	ROMAN NUMERAL TEN
5015	 	<U216A>	ROMAN NUMERAL ELEVEN
5016	<cR>	<U216B>	ROMAN NUMERAL TWELVE
5017	<50R>	<U216C>	ROMAN NUMERAL FIFTY
5018	<100R>	<U216D>	ROMAN NUMERAL ONE HUNDRED
5019	<500R>	<U216E>	ROMAN NUMERAL FIVE HUNDRED
5020	<1000R>	<U216F>	ROMAN NUMERAL ONE THOUSAND
5021	<1r>	<U2170>	SMALL ROMAN NUMERAL ONE
5022	<2r>	<U2171>	SMALL ROMAN NUMERAL TWO
5023	<3r>	<U2172>	SMALL ROMAN NUMERAL THREE
5024	<4r>	<U2173>	SMALL ROMAN NUMERAL FOUR
5025	<5r>	<U2174>	SMALL ROMAN NUMERAL FIVE
5026	<6r>	<U2175>	SMALL ROMAN NUMERAL SIX
5027	<7r>	<U2176>	SMALL ROMAN NUMERAL SEVEN
5028	<8r>	<U2177>	SMALL ROMAN NUMERAL EIGHT
5029	<9r>	<U2178>	SMALL ROMAN NUMERAL NINE
5030	<ar>	<U2179>	SMALL ROMAN NUMERAL TEN
5031	 	<U217A>	SMALL ROMAN NUMERAL ELEVEN
5032	<cr>	<U217B>	SMALL ROMAN NUMERAL TWELVE
5033	<50r>	<U217C>	SMALL ROMAN NUMERAL FIFTY
5034	<100r>	<U217D>	SMALL ROMAN NUMERAL ONE HUNDRED
5035	<500r>	<U217E>	SMALL ROMAN NUMERAL FIVE HUNDRED
5036	<1000r>	<U217F>	SMALL ROMAN NUMERAL ONE THOUSAND
5037	<1000RCD>	<U2180>	ROMAN NUMERAL ONE THOUSAND C D
5038	<5000R>	<U2181>	ROMAN NUMERAL FIVE THOUSAND
5039	<10000R>	<U2182>	ROMAN NUMERAL TEN THOUSAND
5040	<<->	<U2190>	LEFTWARDS ARROW
5041	<-1>	<U2191>	UPWARDS ARROW
5042	<-//>	<U2192>	RIGHTWARDS ARROW
5043	<-v>	<U2193>	DOWNTWARDS ARROW
5044	</->	<U2194>	LEFT RIGHT ARROW
5045	<UD>	<U2195>	UP DOWN ARROW
5046	<!!>	<U2196>	NORTH WEST ARROW
5047	<///>>	<U2197>	NORTH EAST ARROW
5048	<!/>>	<U2198>	SOUTH EAST ARROW
5049	<///>	<U2199>	SOUTH WEST ARROW
5050	<UD->	<U21A8>	UP DOWN ARROW WITH BASE
5051	</>V>	<U21C0>	RIGHTWARDS HARPOON WITH BARB UPWARDS
5052	<==>	<U21D0>	LEFTWARDS DOUBLE ARROW
5053	<=/>	<U21D2>	RIGHTWARDS DOUBLE ARROW
5054	<==>	<U21D4>	LEFT RIGHT DOUBLE ARROW
5055	<FA>	<U2200>	FOR ALL
5056	<dP>	<U2202>	PARTIAL DIFFERENTIAL
5057	<TE>	<U2203>	THERE EXISTS
5058	</>0>	<U2205>	EMPTY SET
5059	<DE>	<U2206>	INCREMENT
5060	<NB>	<U2207>	NABLA
5061	<(->	<U2208>	ELEMENT OF
5062	<->	<U220B>	CONTAINS AS MEMBER
5063	<FP>	<U220E>	END OF PROOF
5064	<*P>	<U220F>	N-ARY PRODUCT
5065	<+Z>	<U2211>	N-ARY SUMMATION
5066	<-2>	<U2212>	MINUS SIGN
5067	<-->	<U2213>	MINUS-OR-PLUS SIGN
5068	<.+>	<U2214>	DOT PLUS
5069	<*->	<U2217>	ASTERISK OPERATOR
5070	<Ob>	<U2218>	RING OPERATOR
5071	<Sb>	<U2219>	BULLET OPERATOR
5072	<RT>	<U221A>	SQUARE ROOT
5073	<0(>	<U221D>	PROPORTIONAL TO
5074	<00>	<U221E>	INFINITY
5075	<-L>	<U221F>	RIGHT ANGLE
5076	<-v>	<U2220>	ANGLE
5077	<PP>	<U2225>	PARALLEL TO
5078	<AN>	<U2227>	LOGICAL AND
5079	<OR>	<U2228>	LOGICAL OR
5080	<(U>	<U2229>	INTERSECTION
5081	<)U>	<U222A>	UNION
5082	<In>	<U222B>	INTEGRAL
5083	<DI>	<U222C>	DOUBLE INTEGRAL
5084	<Io>	<U222E>	CONTOUR INTEGRAL
5085	<.:>	<U2234>	THEREFORE
5086	<.:>	<U2235>	BECAUSE
5087	<:R>	<U2236>	RATIO
5088	<::>	<U2237>	PROPORTION
5089	<?1>	<U223C>	TILDE OPERATOR
5090	<CG>	<U223E>	INVERTED LAZY S

5091	<?->	<U2243>	ASYMPTOTICALLY EQUAL TO
5092	<?=>	<U2245>	APPROXIMATELY EQUAL TO
5093	<?2>	<U2248>	ALMOST EQUAL TO
5094	<=?>	<U224C>	ALL EQUAL TO
5095	<HI>	<U2253>	IMAGE OF OR APPROXIMATELY EQUAL TO
5096	<!->	<U2260>	NOT EQUAL TO
5097	<=3>	<U2261>	IDENTICAL TO
5098	<=<>	<U2264>	LESS-THAN OR EQUAL TO
5099	</>=>	<U2265>	GREATER-THAN OR EQUAL TO
5100	<<*>	<U226A>	MUCH LESS-THAN
5101	<*/>>	<U226B>	MUCH GREATER-THAN
5102	<!<>	<U226E>	NOT LESS-THAN
5103	<!/>>	<U226F>	NOT GREATER-THAN
5104	<(C>	<U2282>	SUBSET OF
5105	<)C>	<U2283>	SUPERSET OF
5106	<(_>	<U2286>	SUBSET OF OR EQUAL TO
5107	<)_>	<U2287>	SUPERSET OF OR EQUAL TO
5108	<0.>	<U2299>	CIRCLED DOT OPERATOR
5109	<02>	<U229A>	CIRCLED RING OPERATOR
5110	<-T>	<U22A5>	UP TACK
5111	<.P>	<U22C5>	DOT OPERATOR
5112	<:3>	<U22EE>	VERTICAL ELLIPSIS
5113	<Eh>	<U2302>	HOUSE
5114	<<7>	<U2308>	LEFT CEILING
5115	</>7>	<U2309>	RIGHT CEILING
5116	<7<>	<U230A>	LEFT FLOOR
5117	<7/>>	<U230B>	RIGHT FLOOR
5118	<NI>	<U2310>	REVERSED NOT SIGN
5119	<(A>	<U2312>	ARC
5120	<TR>	<U2315>	TELEPHONE RECORDER
5121	<88>	<U2318>	PLACE OF INTEREST SIGN
5122	<Iu>	<U2320>	TOP HALF INTEGRAL
5123	<I1>	<U2321>	BOTTOM HALF INTEGRAL
5124	<<//>	<U2329>	LEFT-POINTING ANGLE BRACKET
5125	<//>>	<U232A>	RIGHT-POINTING ANGLE BRACKET
5126	<Vs>	<U2423>	OPEN BOX
5127	<1h>	<U2440>	OCR HOOK
5128	<3h>	<U2441>	OCR CHAIR
5129	<2h>	<U2442>	OCR FORK
5130	<4h>	<U2443>	OCR INVERTED FORK
5131	<1j>	<U2446>	OCR BRANCH BANK IDENTIFICATION
5132	<2j>	<U2447>	OCR AMOUNT OF CHECK
5133	<3j>	<U2448>	OCR DASH
5134	<4j>	<U2449>	OCR CUSTOMER ACCOUNT NUMBER
5135	<1-o>	<U2460>	CIRCLED DIGIT ONE
5136	<2-o>	<U2461>	CIRCLED DIGIT TWO
5137	<3-o>	<U2462>	CIRCLED DIGIT THREE
5138	<4-o>	<U2463>	CIRCLED DIGIT FOUR
5139	<5-o>	<U2464>	CIRCLED DIGIT FIVE
5140	<6-o>	<U2465>	CIRCLED DIGIT SIX
5141	<7-o>	<U2466>	CIRCLED DIGIT SEVEN
5142	<8-o>	<U2467>	CIRCLED DIGIT EIGHT
5143	<9-o>	<U2468>	CIRCLED DIGIT NINE
5144	<10-o>	<U2469>	CIRCLED NUMBER TEN
5145	<11-o>	<U246A>	CIRCLED NUMBER ELEVEN
5146	<12-o>	<U246B>	CIRCLED NUMBER TWELVE
5147	<13-o>	<U246C>	CIRCLED NUMBER THIRTEEN
5148	<14-o>	<U246D>	CIRCLED NUMBER FOURTEEN
5149	<15-o>	<U246E>	CIRCLED NUMBER FIFTEEN
5150	<16-o>	<U246F>	CIRCLED NUMBER SIXTEEN
5151	<17-o>	<U2470>	CIRCLED NUMBER SEVENTEEN
5152	<18-o>	<U2471>	CIRCLED NUMBER EIGHTEEN
5153	<19-o>	<U2472>	CIRCLED NUMBER NINETEEN
5154	<20-o>	<U2473>	CIRCLED NUMBER TWENTY
5155	<(1)>	<U2474>	PARENTHESIZED DIGIT ONE
5156	<(2)>	<U2475>	PARENTHESIZED DIGIT TWO
5157	<(3)>	<U2476>	PARENTHESIZED DIGIT THREE
5158	<(4)>	<U2477>	PARENTHESIZED DIGIT FOUR
5159	<(5)>	<U2478>	PARENTHESIZED DIGIT FIVE
5160	<(6)>	<U2479>	PARENTHESIZED DIGIT SIX
5161	<(7)>	<U247A>	PARENTHESIZED DIGIT SEVEN
5162	<(8)>	<U247B>	PARENTHESIZED DIGIT EIGHT
5163	<(9)>	<U247C>	PARENTHESIZED DIGIT NINE
5164	<(10)>	<U247D>	PARENTHESIZED NUMBER TEN
5165	<(11)>	<U247E>	PARENTHESIZED NUMBER ELEVEN
5166	<(12)>	<U247F>	PARENTHESIZED NUMBER TWELVE
5167	<(13)>	<U2480>	PARENTHESIZED NUMBER THIRTEEN
5168	<(14)>	<U2481>	PARENTHESIZED NUMBER FOURTEEN
5169	<(15)>	<U2482>	PARENTHESIZED NUMBER FIFTEEN
5170	<(16)>	<U2483>	PARENTHESIZED NUMBER SIXTEEN
5171	<(17)>	<U2484>	PARENTHESIZED NUMBER SEVENTEEN
5172	<(18)>	<U2485>	PARENTHESIZED NUMBER EIGHTEEN
5173	<(19)>	<U2486>	PARENTHESIZED NUMBER NINETEEN
5174	<(20)>	<U2487>	PARENTHESIZED NUMBER TWENTY
5175	<1.>	<U2488>	DIGIT ONE FULL STOP
5176	<2.>	<U2489>	DIGIT TWO FULL STOP
5177	<3.>	<U248A>	DIGIT THREE FULL STOP
5178	<4.>	<U248B>	DIGIT FOUR FULL STOP

5 79	<5.>	<U248C>	DIGIT FIVE FULL STOP
5 80	<6.>	<U248D>	DIGIT SIX FULL STOP
5 81	<7.>	<U248E>	DIGIT SEVEN FULL STOP
5 82	<8.>	<U248F>	DIGIT EIGHT FULL STOP
5 83	<9.>	<U2490>	DIGIT NINE FULL STOP
5 84	<10.>	<U2491>	NUMBER TEN FULL STOP
5 85	<11.>	<U2492>	NUMBER ELEVEN FULL STOP
5 86	<12.>	<U2493>	NUMBER TWELVE FULL STOP
5 87	<13.>	<U2494>	NUMBER THIRTEEN FULL STOP
5 88	<14.>	<U2495>	NUMBER FOURTEEN FULL STOP
5 89	<15.>	<U2496>	NUMBER FIFTEEN FULL STOP
5 90	<16.>	<U2497>	NUMBER SIXTEEN FULL STOP
5 91	<17.>	<U2498>	NUMBER SEVENTEEN FULL STOP
5 92	<18.>	<U2499>	NUMBER EIGHTEEN FULL STOP
5 93	<19.>	<U249A>	NUMBER NINETEEN FULL STOP
5 94	<20.>	<U249B>	NUMBER TWENTY FULL STOP
5 95	<(a)>	<U249C>	PARENTHESIZED LATIN SMALL LETTER A
5 96	<(b)>	<U249D>	PARENTHESIZED LATIN SMALL LETTER B
5 97	<(c)>	<U249E>	PARENTHESIZED LATIN SMALL LETTER C
5 98	<(d)>	<U249F>	PARENTHESIZED LATIN SMALL LETTER D
5 99	<(e)>	<U24A0>	PARENTHESIZED LATIN SMALL LETTER E
5 200	<(f)>	<U24A1>	PARENTHESIZED LATIN SMALL LETTER F
5 201	<(g)>	<U24A2>	PARENTHESIZED LATIN SMALL LETTER G
5 202	<(h)>	<U24A3>	PARENTHESIZED LATIN SMALL LETTER H
5 203	<(i)>	<U24A4>	PARENTHESIZED LATIN SMALL LETTER I
5 204	<(j)>	<U24A5>	PARENTHESIZED LATIN SMALL LETTER J
5 205	<(k)>	<U24A6>	PARENTHESIZED LATIN SMALL LETTER K
5 206	<(l)>	<U24A7>	PARENTHESIZED LATIN SMALL LETTER L
5 207	<(m)>	<U24A8>	PARENTHESIZED LATIN SMALL LETTER M
5 208	<(n)>	<U24A9>	PARENTHESIZED LATIN SMALL LETTER N
5 209	<(o)>	<U24AA>	PARENTHESIZED LATIN SMALL LETTER O
5 210	<(p)>	<U24AB>	PARENTHESIZED LATIN SMALL LETTER P
5 211	<(q)>	<U24AC>	PARENTHESIZED LATIN SMALL LETTER Q
5 212	<(r)>	<U24AD>	PARENTHESIZED LATIN SMALL LETTER R
5 213	<(s)>	<U24AE>	PARENTHESIZED LATIN SMALL LETTER S
5 214	<(t)>	<U24AF>	PARENTHESIZED LATIN SMALL LETTER T
5 215	<(u)>	<U24B0>	PARENTHESIZED LATIN SMALL LETTER U
5 216	<(v)>	<U24B1>	PARENTHESIZED LATIN SMALL LETTER V
5 217	<(w)>	<U24B2>	PARENTHESIZED LATIN SMALL LETTER W
5 218	<(x)>	<U24B3>	PARENTHESIZED LATIN SMALL LETTER X
5 219	<(y)>	<U24B4>	PARENTHESIZED LATIN SMALL LETTER Y
5 220	<(z)>	<U24B5>	PARENTHESIZED LATIN SMALL LETTER Z
5 221	<A-o>	<U24B6>	CIRCLED LATIN CAPITAL LETTER A
5 222	<B-o>	<U24B7>	CIRCLED LATIN CAPITAL LETTER B
5 223	<C-o>	<U24B8>	CIRCLED LATIN CAPITAL LETTER C
5 224	<D-o>	<U24B9>	CIRCLED LATIN CAPITAL LETTER D
5 225	<E-o>	<U24BA>	CIRCLED LATIN CAPITAL LETTER E
5 226	<F-o>	<U24BB>	CIRCLED LATIN CAPITAL LETTER F
5 227	<G-o>	<U24BC>	CIRCLED LATIN CAPITAL LETTER G
5 228	<H-o>	<U24BD>	CIRCLED LATIN CAPITAL LETTER H
5 229	<I-o>	<U24BE>	CIRCLED LATIN CAPITAL LETTER I
5 230	<J-o>	<U24BF>	CIRCLED LATIN CAPITAL LETTER J
5 231	<K-o>	<U24C0>	CIRCLED LATIN CAPITAL LETTER K
5 232	<L-o>	<U24C1>	CIRCLED LATIN CAPITAL LETTER L
5 233	<M-o>	<U24C2>	CIRCLED LATIN CAPITAL LETTER M
5 234	<N-o>	<U24C3>	CIRCLED LATIN CAPITAL LETTER N
5 235	<O-o>	<U24C4>	CIRCLED LATIN CAPITAL LETTER O
5 236	<P-o>	<U24C5>	CIRCLED LATIN CAPITAL LETTER P
5 237	<Q-o>	<U24C6>	CIRCLED LATIN CAPITAL LETTER Q
5 238	<R-o>	<U24C7>	CIRCLED LATIN CAPITAL LETTER R
5 239	<S-o>	<U24C8>	CIRCLED LATIN CAPITAL LETTER S
5 240	<T-o>	<U24C9>	CIRCLED LATIN CAPITAL LETTER T
5 241	<U-o>	<U24CA>	CIRCLED LATIN CAPITAL LETTER U
5 242	<V-o>	<U24CB>	CIRCLED LATIN CAPITAL LETTER V
5 243	<W-o>	<U24CC>	CIRCLED LATIN CAPITAL LETTER W
5 244	<X-o>	<U24CD>	CIRCLED LATIN CAPITAL LETTER X
5 245	<Y-o>	<U24CE>	CIRCLED LATIN CAPITAL LETTER Y
5 246	<Z-o>	<U24CF>	CIRCLED LATIN CAPITAL LETTER Z
5 247	<a-o>	<U24D0>	CIRCLED LATIN SMALL LETTER A
5 248	<b-o>	<U24D1>	CIRCLED LATIN SMALL LETTER B
5 249	<c-o>	<U24D2>	CIRCLED LATIN SMALL LETTER C
5 250	<d-o>	<U24D3>	CIRCLED LATIN SMALL LETTER D
5 251	<e-o>	<U24D4>	CIRCLED LATIN SMALL LETTER E
5 252	<f-o>	<U24D5>	CIRCLED LATIN SMALL LETTER F
5 253	<g-o>	<U24D6>	CIRCLED LATIN SMALL LETTER G
5 254	<h-o>	<U24D7>	CIRCLED LATIN SMALL LETTER H
5 255	<i-o>	<U24D8>	CIRCLED LATIN SMALL LETTER I
5 256	<j-o>	<U24D9>	CIRCLED LATIN SMALL LETTER J
5 257	<k-o>	<U24DA>	CIRCLED LATIN SMALL LETTER K
5 258	<l-o>	<U24DB>	CIRCLED LATIN SMALL LETTER L
5 259	<m-o>	<U24DC>	CIRCLED LATIN SMALL LETTER M
5 260	<n-o>	<U24DD>	CIRCLED LATIN SMALL LETTER N
5 261	<o-o>	<U24DE>	CIRCLED LATIN SMALL LETTER O
5 262	<p-o>	<U24DF>	CIRCLED LATIN SMALL LETTER P
5 263	<q-o>	<U24E0>	CIRCLED LATIN SMALL LETTER Q
5 264	<r-o>	<U24E1>	CIRCLED LATIN SMALL LETTER R
5 265	<s-o>	<U24E2>	CIRCLED LATIN SMALL LETTER S
5 266	<t-o>	<U24E3>	CIRCLED LATIN SMALL LETTER T

5267	<u-o>	<U24E4>	CIRCLED LATIN SMALL LETTER U
5268	<v-o>	<U24E5>	CIRCLED LATIN SMALL LETTER V
5269	<w-o>	<U24E6>	CIRCLED LATIN SMALL LETTER W
5270	<x-o>	<U24E7>	CIRCLED LATIN SMALL LETTER X
5271	<y-o>	<U24E8>	CIRCLED LATIN SMALL LETTER Y
5272	<z-o>	<U24E9>	CIRCLED LATIN SMALL LETTER Z
5273	<0-o>	<U24EA>	CIRCLED DIGIT ZERO
5274	<hh>	<U2500>	BOX DRAWINGS LIGHT HORIZONTAL
5275	<HH>	<U2501>	BOX DRAWINGS HEAVY HORIZONTAL
5276	<vv>	<U2502>	BOX DRAWINGS LIGHT VERTICAL
5277	<VV>	<U2503>	BOX DRAWINGS HEAVY VERTICAL
5278	<3->	<U2504>	BOX DRAWINGS LIGHT TRIPLE DASH HORIZONTAL
5279	<3_>	<U2505>	BOX DRAWINGS HEAVY TRIPLE DASH HORIZONTAL
5280	<3>	<U2506>	BOX DRAWINGS LIGHT TRIPLE DASH VERTICAL
5281	<3//>	<U2507>	BOX DRAWINGS HEAVY TRIPLE DASH VERTICAL
5282	<4->	<U2508>	BOX DRAWINGS LIGHT QUADRUPLE DASH HORIZONTAL
5283	<4_>	<U2509>	BOX DRAWINGS HEAVY QUADRUPLE DASH HORIZONTAL
5284	<4>	<U250A>	BOX DRAWINGS LIGHT QUADRUPLE DASH VERTICAL
5285	<4//>	<U250B>	BOX DRAWINGS HEAVY QUADRUPLE DASH VERTICAL
5286	<dR>	<U250C>	BOX DRAWINGS LIGHT DOWN AND RIGHT
5287	<dR->	<U250D>	BOX DRAWINGS DOWN LIGHT AND RIGHT HEAVY
5288	<Dr->	<U250E>	BOX DRAWINGS DOWN HEAVY AND RIGHT LIGHT
5289	<DR->	<U250F>	BOX DRAWINGS HEAVY DOWN AND RIGHT
5290	<dL>	<U2510>	BOX DRAWINGS LIGHT DOWN AND LEFT
5291	<dL->	<U2511>	BOX DRAWINGS DOWN LIGHT AND LEFT HEAVY
5292	<DL->	<U2512>	BOX DRAWINGS DOWN HEAVY AND LEFT LIGHT
5293	<LD->	<U2513>	BOX DRAWINGS HEAVY DOWN AND LEFT
5294	<uR>	<U2514>	BOX DRAWINGS LIGHT UP AND RIGHT
5295	<uR->	<U2515>	BOX DRAWINGS UP LIGHT AND RIGHT HEAVY
5296	<Uz->	<U2516>	BOX DRAWINGS UP HEAVY AND RIGHT LIGHT
5297	<UR->	<U2517>	BOX DRAWINGS HEAVY UP AND RIGHT
5298	<uL>	<U2518>	BOX DRAWINGS LIGHT UP AND LEFT
5299	<uL->	<U2519>	BOX DRAWINGS UP LIGHT AND LEFT HEAVY
5300	<U1->	<U251A>	BOX DRAWINGS UP HEAVY AND LEFT LIGHT
5301	<UL->	<U251B>	BOX DRAWINGS HEAVY UP AND LEFT
5302	<vz>	<U251C>	BOX DRAWINGS LIGHT VERTICAL AND RIGHT
5303	<vR->	<U251D>	BOX DRAWINGS VERTICAL LIGHT AND RIGHT HEAVY
5304	<Udr>	<U251E>	BOX DRAWINGS UP HEAVY AND RIGHT DOWN LIGHT
5305	<uDz>	<U251F>	BOX DRAWINGS DOWN HEAVY AND RIGHT UP LIGHT
5306	<Vr->	<U2520>	BOX DRAWINGS VERTICAL HEAVY AND RIGHT LIGHT
5307	<UdR>	<U2521>	BOX DRAWINGS DOWN LIGHT AND RIGHT UP HEAVY
5308	<uDR>	<U2522>	BOX DRAWINGS UP LIGHT AND RIGHT DOWN HEAVY
5309	<VR->	<U2523>	BOX DRAWINGS HEAVY VERTICAL AND RIGHT
5310	<v1>	<U2524>	BOX DRAWINGS LIGHT VERTICAL AND LEFT
5311	<VL->	<U2525>	BOX DRAWINGS VERTICAL LIGHT AND LEFT HEAVY
5312	<Ud1>	<U2526>	BOX DRAWINGS UP HEAVY AND LEFT DOWN LIGHT
5313	<uD1>	<U2527>	BOX DRAWINGS DOWN HEAVY AND LEFT UP LIGHT
5314	<V1->	<U2528>	BOX DRAWINGS VERTICAL HEAVY AND LEFT LIGHT
5315	<UdL>	<U2529>	BOX DRAWINGS DOWN LIGHT AND LEFT UP HEAVY
5316	<uDL>	<U252A>	BOX DRAWINGS UP LIGHT AND LEFT DOWN HEAVY
5317	<VL->	<U252B>	BOX DRAWINGS HEAVY VERTICAL AND LEFT
5318	<dh>	<U252C>	BOX DRAWINGS LIGHT DOWN AND HORIZONTAL
5319	<dLr>	<U252D>	BOX DRAWINGS LEFT HEAVY AND RIGHT DOWN LIGHT
5320	<dLR>	<U252E>	BOX DRAWINGS RIGHT HEAVY AND LEFT DOWN LIGHT
5321	<dH->	<U252F>	BOX DRAWINGS DOWN LIGHT AND HORIZONTAL HEAVY
5322	<Dh->	<U2530>	BOX DRAWINGS DOWN HEAVY AND HORIZONTAL LIGHT
5323	<DLr>	<U2531>	BOX DRAWINGS RIGHT LIGHT AND LEFT DOWN HEAVY
5324	<DLR>	<U2532>	BOX DRAWINGS LEFT LIGHT AND RIGHT DOWN HEAVY
5325	<DH->	<U2533>	BOX DRAWINGS HEAVY DOWN AND HORIZONTAL
5326	<uh>	<U2534>	BOX DRAWINGS LIGHT UP AND HORIZONTAL
5327	<ULr>	<U2535>	BOX DRAWINGS LEFT HEAVY AND RIGHT UP LIGHT
5328	<u1R>	<U2536>	BOX DRAWINGS RIGHT HEAVY AND LEFT UP LIGHT
5329	<uH->	<U2537>	BOX DRAWINGS UP LIGHT AND HORIZONTAL HEAVY
5330	<Uh->	<U2538>	BOX DRAWINGS UP HEAVY AND HORIZONTAL LIGHT
5331	<ULr>	<U2539>	BOX DRAWINGS RIGHT LIGHT AND LEFT UP HEAVY
5332	<U1R>	<U253A>	BOX DRAWINGS LEFT LIGHT AND RIGHT UP HEAVY
5333	<UH->	<U253B>	BOX DRAWINGS HEAVY UP AND HORIZONTAL
5334	<vh>	<U253C>	BOX DRAWINGS LIGHT VERTICAL AND HORIZONTAL
5335	<VLr>	<U253D>	BOX DRAWINGS LEFT HEAVY AND RIGHT VERTICAL LIGHT
5336	<v1R>	<U253E>	BOX DRAWINGS RIGHT HEAVY AND LEFT VERTICAL LIGHT
5337	<vH->	<U253F>	BOX DRAWINGS VERTICAL LIGHT AND HORIZONTAL HEAVY
5338	<Udh>	<U2540>	BOX DRAWINGS UP HEAVY AND DOWN HORIZONTAL LIGHT
5339	<uDh>	<U2541>	BOX DRAWINGS DOWN HEAVY AND UP HORIZONTAL LIGHT
5340	<vh->	<U2542>	BOX DRAWINGS VERTICAL HEAVY AND HORIZONTAL LIGHT
5341	<UdLr>	<U2543>	BOX DRAWINGS LEFT UP HEAVY AND RIGHT DOWN LIGHT
5342	<udLR>	<U2544>	BOX DRAWINGS RIGHT UP HEAVY AND LEFT DOWN LIGHT
5343	<uDLr>	<U2545>	BOX DRAWINGS LEFT DOWN HEAVY AND RIGHT UP LIGHT
5344	<uDLR>	<U2546>	BOX DRAWINGS RIGHT DOWN HEAVY AND LEFT UP LIGHT
5345	<UdH>	<U2547>	BOX DRAWINGS DOWN LIGHT AND UP HORIZONTAL HEAVY
5346	<uDH>	<U2548>	BOX DRAWINGS UP LIGHT AND DOWN HORIZONTAL HEAVY
5347	<VLr>	<U2549>	BOX DRAWINGS RIGHT LIGHT AND LEFT VERTICAL HEAVY
5348	<V1R>	<U254A>	BOX DRAWINGS LEFT LIGHT AND RIGHT VERTICAL HEAVY
5349	<VH->	<U254B>	BOX DRAWINGS HEAVY VERTICAL AND HORIZONTAL
5350	<HH>	<U2550>	BOX DRAWINGS DOUBLE HORIZONTAL
5351	<VV>	<U2551>	BOX DRAWINGS DOUBLE VERTICAL
5352	<dR>	<U2552>	BOX DRAWINGS DOWN SINGLE AND RIGHT DOUBLE
5353	<Dr>	<U2553>	BOX DRAWINGS DOWN DOUBLE AND RIGHT SINGLE
5354	<DR>	<U2554>	BOX DRAWINGS DOUBLE DOWN AND RIGHT

5355	<dL>	<U2555>	BOX DRAWINGS DOWN SINGLE AND LEFT DOUBLE
5356	<D1>	<U2556>	BOX DRAWINGS DOWN DOUBLE AND LEFT SINGLE
5357	<LD>	<U2557>	BOX DRAWINGS DOUBLE DOWN AND LEFT
5358	<uR>	<U2558>	BOX DRAWINGS UP SINGLE AND RIGHT DOUBLE
5359	<Ur>	<U2559>	BOX DRAWINGS UP DOUBLE AND RIGHT SINGLE
5360	<UR>	<U255A>	BOX DRAWINGS DOUBLE UP AND RIGHT
5361	<uL>	<U255B>	BOX DRAWINGS UP SINGLE AND LEFT DOUBLE
5362	<U1>	<U255C>	BOX DRAWINGS UP DOUBLE AND LEFT SINGLE
5363	<UL>	<U255D>	BOX DRAWINGS DOUBLE UP AND LEFT
5364	<vR>	<U255E>	BOX DRAWINGS VERTICAL SINGLE AND RIGHT DOUBLE
5365	<Vr>	<U255F>	BOX DRAWINGS VERTICAL DOUBLE AND RIGHT SINGLE
5366	<VR>	<U2560>	BOX DRAWINGS DOUBLE VERTICAL AND RIGHT
5367	<vL>	<U2561>	BOX DRAWINGS VERTICAL SINGLE AND LEFT DOUBLE
5368	<V1>	<U2562>	BOX DRAWINGS VERTICAL DOUBLE AND LEFT SINGLE
5369	<VL>	<U2563>	BOX DRAWINGS DOUBLE VERTICAL AND LEFT
5370	<dH>	<U2564>	BOX DRAWINGS DOWN SINGLE AND HORIZONTAL DOUBLE
5371	<Dh>	<U2565>	BOX DRAWINGS DOWN DOUBLE AND HORIZONTAL SINGLE
5372	<DH>	<U2566>	BOX DRAWINGS DOUBLE DOWN AND HORIZONTAL
5373	<uH>	<U2567>	BOX DRAWINGS UP SINGLE AND HORIZONTAL DOUBLE
5374	<Uh>	<U2568>	BOX DRAWINGS UP DOUBLE AND HORIZONTAL SINGLE
5375	<UH>	<U2569>	BOX DRAWINGS DOUBLE UP AND HORIZONTAL
5376	<vH>	<U256A>	BOX DRAWINGS VERTICAL SINGLE AND HORIZONTAL DOUBLE
5377	<Vh>	<U256B>	BOX DRAWINGS VERTICAL DOUBLE AND HORIZONTAL SINGLE
5378	<VH>	<U256C>	BOX DRAWINGS DOUBLE VERTICAL AND HORIZONTAL
5379	<FD>	<U2571>	BOX DRAWINGS LIGHT DIAGONAL UPPER RIGHT TO LOWER LEFT
5380	<BD>	<U2572>	BOX DRAWINGS LIGHT DIAGONAL UPPER LEFT TO LOWER RIGHT
5381	<TB>	<U2580>	UPPER HALF BLOCK
5382	<LB>	<U2584>	LOWER HALF BLOCK
5383	<FB>	<U2588>	FULL BLOCK
5384	<1B>	<U258C>	LEFT HALF BLOCK
5385	<RB>	<U2590>	RIGHT HALF BLOCK
5386	<.S>	<U2591>	LIGHT SHADE
5387	<:S>	<U2592>	MEDIUM SHADE
5388	<?S>	<U2593>	DARK SHADE
5389	<fS>	<U25A0>	BLACK SQUARE
5390	<oS>	<U25A1>	WHITE SQUARE
5391	<rO>	<U25A2>	WHITE SQUARE WITH ROUNDED CORNERS
5392	<Rx>	<U25A3>	WHITE SQUARE CONTAINING BLACK SMALL SQUARE
5393	<RF>	<U25A4>	SQUARE WITH HORIZONTAL FILL
5394	<RY>	<U25A5>	SQUARE WITH VERTICAL FILL
5395	<RH>	<U25A6>	SQUARE WITH ORTHOGONAL CROSHATCH FILL
5396	<RZ>	<U25A7>	SQUARE WITH UPPER LEFT TO LOWER RIGHT FILL
5397	<RK>	<U25A8>	SQUARE WITH UPPER RIGHT TO LOWER LEFT FILL
5398	<RX>	<U25A9>	SQUARE WITH DIAGONAL CROSHATCH FILL
5399	<sB>	<U25AA>	BLACK SMALL SQUARE
5400	<SR>	<U25AC>	BLACK RECTANGLE
5401	<Or>	<U25AD>	WHITE RECTANGLE
5402	<UT>	<U25B2>	BLACK UP-POINTING TRIANGLE
5403	<uT>	<U25B3>	WHITE UP-POINTING TRIANGLE
5404	<Tr>	<U25B7>	WHITE RIGHT-POINTING TRIANGLE
5405	<PR>	<U25BA>	BLACK RIGHT-POINTING POINTER
5406	<Dt>	<U25BC>	BLACK DOWN-POINTING TRIANGLE
5407	<dT>	<U25BD>	WHITE DOWN-POINTING TRIANGLE
5408	<Tl>	<U25C1>	WHITE LEFT-POINTING TRIANGLE
5409	<PL>	<U25C4>	BLACK LEFT-POINTING POINTER
5410	<Db>	<U25C6>	BLACK DIAMOND
5411	<Dw>	<U25C7>	WHITE DIAMOND
5412	<LZ>	<U25CA>	LOZENGE
5413	<0m>	<U25CB>	WHITE CIRCLE
5414	<0o>	<U25CE>	BULLSEYE
5415	<0M>	<U25CF>	BLACK CIRCLE
5416	<0L>	<U25D0>	CIRCLE WITH LEFT HALF BLACK
5417	<0R>	<U25D1>	CIRCLE WITH RIGHT HALF BLACK
5418	<Sn>	<U25D8>	INVERSE BULLET
5419	<IC>	<U25D9>	INVERSE WHITE CIRCLE
5420	<Fd>	<U25E2>	BLACK LOWER RIGHT TRIANGLE
5421	<Bd>	<U25E3>	BLACK LOWER LEFT TRIANGLE
5422	<Ci>	<U25EF>	LARGE CIRCLE
5423	<*2>	<U2605>	BLACK STAR
5424	<*1>	<U2606>	WHITE STAR
5425	<TEL>	<U260E>	BLACK TELEPHONE
5426	<tel>	<U260F>	WHITE TELEPHONE
5427	<<H>	<U261C>	WHITE LEFT POINTING INDEX
5428	</>H>	<U261E>	WHITE RIGHT POINTING INDEX
5429	<0u>	<U263A>	WHITE SMILING FACE
5430	<0U>	<U263B>	BLACK SMILING FACE
5431	<SU>	<U263C>	WHITE SUN WITH RAYS
5432	<Fm>	<U2640>	FEMALE SIGN
5433	<M1>	<U2642>	MALE SIGN
5434	<cS>	<U2660>	BLACK SPADE SUIT
5435	<cH>	<U2661>	WHITE HEART SUIT
5436	<cD>	<U2662>	WHITE DIAMOND SUIT
5437	<cC>	<U2663>	BLACK CLUB SUIT
5438	<cS->	<U2664>	WHITE SPADE SUIT
5439	<cH->	<U2665>	BLACK HEART SUIT
5440	<cD->	<U2666>	BLACK DIAMOND SUIT
5441	<cC->	<U2667>	WHITE CLUB SUIT
5442	<Md>	<U2669>	QUARTER NOTE

5443	<M8>	<U266A>	EIGHTH NOTE
5444	<M2>	<U266B>	BEAMED EIGHTH NOTES
5445	<M16>	<U266C>	BEAMED SIXTEENTH NOTES
5446	<Mb>	<U266D>	MUSIC FLAT SIGN
5447	<Mx>	<U266E>	MUSIC NATURAL SIGN
5448	<MX>	<U266F>	MUSIC SHARP SIGN
5449	<OK>	<U2713>	CHECK MARK
5450	<XX>	<U2717>	BALLOT X
5451	<-X>	<U2720>	MALTESE CROSS
5452	<IS>	<U3000>	IDEOGRAPHIC SPACE
5453	<,_>	<U3001>	IDEOGRAPHIC COMMA
5454	<_.>	<U3002>	IDEOGRAPHIC FULL STOP
5455	<+>	<U3003>	DITTO MARK
5456	<JIS>	<U3004>	JAPANESE INDUSTRIAL STANDARD SYMBOL
5457	<*_>	<U3005>	IDEOGRAPHIC ITERATION MARK
5458	<;_>	<U3006>	IDEOGRAPHIC CLOSING MARK
5459	<0_>	<U3007>	IDEOGRAPHIC NUMBER ZERO
5460	<>+>	<U300A>	LEFT DOUBLE ANGLE BRACKET
5461	</>+>	<U300B>	RIGHT DOUBLE ANGLE BRACKET
5462	<<_>	<U300C>	LEFT CORNER BRACKET
5463	</>'>	<U300D>	RIGHT CORNER BRACKET
5464	<<_>	<U300E>	LEFT WHITE CORNER BRACKET
5465	</>">	<U300F>	RIGHT WHITE CORNER BRACKET
5466	<(>	<U3010>	LEFT BLACK LENTICULAR BRACKET
5467	<)">	<U3011>	RIGHT BLACK LENTICULAR BRACKET
5468	<=T>	<U3012>	POSTAL MARK
5469	<=_>	<U3013>	GETA MARK
5470	<(>'	<U3014>	LEFT TORTOISE SHELL BRACKET
5471	<)">	<U3015>	RIGHT TORTOISE SHELL BRACKET
5472	<I>	<U3016>	LEFT WHITE LENTICULAR BRACKET
5473	<)I>	<U3017>	RIGHT WHITE LENTICULAR BRACKET
5474	<?->	<U301C>	WAVE DASH
5475	<=T:>	<U3020>	POSTAL MARK FACE
5476	<A5>	<U3041>	HIRAGANA LETTER SMALL A
5477	<a5>	<U3042>	HIRAGANA LETTER A
5478	<i5>	<U3043>	HIRAGANA LETTER SMALL I
5479	<i5>	<U3044>	HIRAGANA LETTER I
5480	<U5>	<U3045>	HIRAGANA LETTER SMALL U
5481	<u5>	<U3046>	HIRAGANA LETTER U
5482	<E5>	<U3047>	HIRAGANA LETTER SMALL E
5483	<e5>	<U3048>	HIRAGANA LETTER E
5484	<O5>	<U3049>	HIRAGANA LETTER SMALL O
5485	<o5>	<U304A>	HIRAGANA LETTER O
5486	<ka>	<U304B>	HIRAGANA LETTER KA
5487	<ga>	<U304C>	HIRAGANA LETTER GA
5488	<ki>	<U304D>	HIRAGANA LETTER KI
5489	<gi>	<U304E>	HIRAGANA LETTER GI
5490	<ku>	<U304F>	HIRAGANA LETTER KU
5491	<gu>	<U3050>	HIRAGANA LETTER GU
5492	<ke>	<U3051>	HIRAGANA LETTER KE
5493	<ge>	<U3052>	HIRAGANA LETTER GE
5494	<ko>	<U3053>	HIRAGANA LETTER KO
5495	<go>	<U3054>	HIRAGANA LETTER GO
5496	<sa>	<U3055>	HIRAGANA LETTER SA
5497	<za>	<U3056>	HIRAGANA LETTER ZA
5498	<si>	<U3057>	HIRAGANA LETTER SI
5499	<zii>	<U3058>	HIRAGANA LETTER ZI
5500	<su>	<U3059>	HIRAGANA LETTER SU
5501	<zu>	<U305A>	HIRAGANA LETTER ZU
5502	<se>	<U305B>	HIRAGANA LETTER SE
5503	<ze>	<U305C>	HIRAGANA LETTER ZE
5504	<so>	<U305D>	HIRAGANA LETTER SO
5505	<zo>	<U305E>	HIRAGANA LETTER ZO
5506	<ta>	<U305F>	HIRAGANA LETTER TA
5507	<da>	<U3060>	HIRAGANA LETTER DA
5508	<ti>	<U3061>	HIRAGANA LETTER TI
5509	<di>	<U3062>	HIRAGANA LETTER DI
5510	<tU>	<U3063>	HIRAGANA LETTER SMALL TU
5511	<tu>	<U3064>	HIRAGANA LETTER TU
5512	<du>	<U3065>	HIRAGANA LETTER DU
5513	<te>	<U3066>	HIRAGANA LETTER TE
5514	<de>	<U3067>	HIRAGANA LETTER DE
5515	<to>	<U3068>	HIRAGANA LETTER TO
5516	<do>	<U3069>	HIRAGANA LETTER DO
5517	<na>	<U306A>	HIRAGANA LETTER NA
5518	<ni>	<U306B>	HIRAGANA LETTER NI
5519	<nu>	<U306C>	HIRAGANA LETTER NU
5520	<ne>	<U306D>	HIRAGANA LETTER NE
5521	<no>	<U306E>	HIRAGANA LETTER NO
5522	<ha>	<U306F>	HIRAGANA LETTER HA
5523	<ba>	<U3070>	HIRAGANA LETTER BA
5524	<pa>	<U3071>	HIRAGANA LETTER PA
5525	<hi>	<U3072>	HIRAGANA LETTER HI
5526	<bi>	<U3073>	HIRAGANA LETTER BI
5527	<pi>	<U3074>	HIRAGANA LETTER PI
5528	<hu>	<U3075>	HIRAGANA LETTER HU
5529	<bu>	<U3076>	HIRAGANA LETTER BU
5530	<pu>	<U3077>	HIRAGANA LETTER PU

5531	<he>	<U3078>	HIRAGANA LETTER HE
5532	<be>	<U3079>	HIRAGANA LETTER BE
5533	<pe>	<U307A>	HIRAGANA LETTER PE
5534	<ho>	<U307B>	HIRAGANA LETTER HO
5535	<bo>	<U307C>	HIRAGANA LETTER BO
5536	<po>	<U307D>	HIRAGANA LETTER PO
5537	<ma>	<U307E>	HIRAGANA LETTER MA
5538	<mi>	<U307F>	HIRAGANA LETTER MI
5539	<mu>	<U3080>	HIRAGANA LETTER MU
5540	<me>	<U3081>	HIRAGANA LETTER ME
5541	<mo>	<U3082>	HIRAGANA LETTER MO
5542	<yA>	<U3083>	HIRAGANA LETTER SMALL YA
5543	<ya>	<U3084>	HIRAGANA LETTER YA
5544	<yU>	<U3085>	HIRAGANA LETTER SMALL YU
5545	<yu>	<U3086>	HIRAGANA LETTER YU
5546	<yO>	<U3087>	HIRAGANA LETTER SMALL YO
5547	<yo>	<U3088>	HIRAGANA LETTER YO
5548	<ra>	<U3089>	HIRAGANA LETTER RA
5549	<ri>	<U308A>	HIRAGANA LETTER RI
5550	<ru>	<U308B>	HIRAGANA LETTER RU
5551	<re>	<U308C>	HIRAGANA LETTER RE
5552	<ro>	<U308D>	HIRAGANA LETTER RO
5553	<wA>	<U308E>	HIRAGANA LETTER SMALL WA
5554	<wa>	<U308F>	HIRAGANA LETTER WA
5555	<wi>	<U3090>	HIRAGANA LETTER WI
5556	<we>	<U3091>	HIRAGANA LETTER WE
5557	<wo>	<U3092>	HIRAGANA LETTER WO
5558	<n>	<U3093>	HIRAGANA LETTER N
5559	<vu>	<U3094>	HIRAGANA LETTER VU
5560	<"5">	<U309B>	KATAKANA-HIRAGANA VOICED SOUND MARK
5561	<05>	<U309C>	KATAKANA-HIRAGANA SEMI-VOICED SOUND MARK
5562	<*5>	<U309D>	HIRAGANA ITERATION MARK
5563	<+5>	<U309E>	HIRAGANA VOICED ITERATION MARK
5564	<a6>	<U30A1>	KATAKANA LETTER SMALL A
5565	<A6>	<U30A2>	KATAKANA LETTER A
5566	<i6>	<U30A3>	KATAKANA LETTER SMALL I
5567	<I6>	<U30A4>	KATAKANA LETTER I
5568	<u6>	<U30A5>	KATAKANA LETTER SMALL U
5569	<U6>	<U30A6>	KATAKANA LETTER U
5570	<e6>	<U30A7>	KATAKANA LETTER SMALL E
5571	<E6>	<U30A8>	KATAKANA LETTER E
5572	<o6>	<U30A9>	KATAKANA LETTER SMALL O
5573	<O6>	<U30AA>	KATAKANA LETTER O
5574	<Ka>	<U30AB>	KATAKANA LETTER KA
5575	<Ga>	<U30AC>	KATAKANA LETTER GA
5576	<Ki>	<U30AD>	KATAKANA LETTER KI
5577	<Gi>	<U30AE>	KATAKANA LETTER GI
5578	<Ku>	<U30AF>	KATAKANA LETTER KU
5579	<Gu>	<U30B0>	KATAKANA LETTER GU
5580	<Ke>	<U30B1>	KATAKANA LETTER KE
5581	<Ge>	<U30B2>	KATAKANA LETTER GE
5582	<Ko>	<U30B3>	KATAKANA LETTER KO
5583	<Go>	<U30B4>	KATAKANA LETTER GO
5584	<Sa>	<U30B5>	KATAKANA LETTER SA
5585	<Za>	<U30B6>	KATAKANA LETTER ZA
5586	<Si>	<U30B7>	KATAKANA LETTER SI
5587	<Zi>	<U30B8>	KATAKANA LETTER ZI
5588	<Su>	<U30B9>	KATAKANA LETTER SU
5589	<Zu>	<U30BA>	KATAKANA LETTER ZU
5590	<Se>	<U30BB>	KATAKANA LETTER SE
5591	<Ze>	<U30BC>	KATAKANA LETTER ZE
5592	<So>	<U30BD>	KATAKANA LETTER SO
5593	<Zo>	<U30BE>	KATAKANA LETTER ZO
5594	<Ta>	<U30BF>	KATAKANA LETTER TA
5595	<Da>	<U30C0>	KATAKANA LETTER DA
5596	<Ti>	<U30C1>	KATAKANA LETTER TI
5597	<Di>	<U30C2>	KATAKANA LETTER DI
5598	<TU>	<U30C3>	KATAKANA LETTER SMALL TU
5599	<Tu>	<U30C4>	KATAKANA LETTER TU
5600	<Du>	<U30C5>	KATAKANA LETTER DU
5601	<Te>	<U30C6>	KATAKANA LETTER TE
5602	<De>	<U30C7>	KATAKANA LETTER DE
5603	<To>	<U30C8>	KATAKANA LETTER TO
5604	<Do>	<U30C9>	KATAKANA LETTER DO
5605	<Na>	<U30CA>	KATAKANA LETTER NA
5606	<Ni>	<U30CB>	KATAKANA LETTER NI
5607	<Nu>	<U30CC>	KATAKANA LETTER NU
5608	<Ne>	<U30CD>	KATAKANA LETTER NE
5609	<No>	<U30CE>	KATAKANA LETTER NO
5610	<Ha>	<U30CF>	KATAKANA LETTER HA
5611	<Ba>	<U30D0>	KATAKANA LETTER BA
5612	<Pa>	<U30D1>	KATAKANA LETTER PA
5613	<Hi>	<U30D2>	KATAKANA LETTER HI
5614	<Bi>	<U30D3>	KATAKANA LETTER BI
5615	<Pi>	<U30D4>	KATAKANA LETTER PI
5616	<Hu>	<U30D5>	KATAKANA LETTER HU
5617	<Bu>	<U30D6>	KATAKANA LETTER BU
5618	<Pu>	<U30D7>	KATAKANA LETTER PU

5619	<He>	<U30D8>	KATAKANA LETTER HE
5620	<Be>	<U30D9>	KATAKANA LETTER BE
5621	<Pe>	<U30DA>	KATAKANA LETTER PE
5622	<Ho>	<U30DB>	KATAKANA LETTER HO
5623	<Bo>	<U30DC>	KATAKANA LETTER BO
5624	<Po>	<U30DD>	KATAKANA LETTER PO
5625	<Ma>	<U30DE>	KATAKANA LETTER MA
5626	<Mi>	<U30DF>	KATAKANA LETTER MI
5627	<Mu>	<U30E0>	KATAKANA LETTER MU
5628	<Me>	<U30E1>	KATAKANA LETTER ME
5629	<Mo>	<U30E2>	KATAKANA LETTER MO
5630	<YA>	<U30E3>	KATAKANA LETTER SMALL YA
5631	<Ya>	<U30E4>	KATAKANA LETTER YA
5632	<YU>	<U30E5>	KATAKANA LETTER SMALL YU
5633	<Yu>	<U30E6>	KATAKANA LETTER YU
5634	<YO>	<U30E7>	KATAKANA LETTER SMALL YO
5635	<Yo>	<U30E8>	KATAKANA LETTER YO
5636	<Ra>	<U30E9>	KATAKANA LETTER RA
5637	<Ri>	<U30EA>	KATAKANA LETTER RI
5638	<Ru>	<U30EB>	KATAKANA LETTER RU
5639	<Re>	<U30EC>	KATAKANA LETTER RE
5640	<Ro>	<U30ED>	KATAKANA LETTER RO
5641	<WA>	<U30EE>	KATAKANA LETTER SMALL WA
5642	<Wa>	<U30EF>	KATAKANA LETTER WA
5643	<Wi>	<U30F0>	KATAKANA LETTER WI
5644	<We>	<U30F1>	KATAKANA LETTER WE
5645	<Wo>	<U30F2>	KATAKANA LETTER WO
5646	<N6>	<U30F3>	KATAKANA LETTER N
5647	<Vu>	<U30F4>	KATAKANA LETTER VU
5648	<KA>	<U30F5>	KATAKANA LETTER SMALL KA
5649	<KE>	<U30F6>	KATAKANA LETTER SMALL KE
5650	<Va>	<U30F7>	KATAKANA LETTER VA
5651	<Vi>	<U30F8>	KATAKANA LETTER VI
5652	<Ve>	<U30F9>	KATAKANA LETTER VE
5653	<Vo>	<U30FA>	KATAKANA LETTER VO
5654	<.6>	<U30FB>	KATAKANA MIDDLE DOT
5655	<-6>	<U30FC>	KATAKANA-HIRAGANA PROLONGED SOUND MARK
5656	<*6>	<U30FD>	KATAKANA ITERATION MARK
5657	<+6>	<U30FE>	KATAKANA VOICED ITERATION MARK
5658	<b4>	<U3105>	BOPOMOFO LETTER B
5659	<p4>	<U3106>	BOPOMOFO LETTER P
5660	<m4>	<U3107>	BOPOMOFO LETTER M
5661	<f4>	<U3108>	BOPOMOFO LETTER F
5662	<d4>	<U3109>	BOPOMOFO LETTER D
5663	<t4>	<U310A>	BOPOMOFO LETTER T
5664	<n4>	<U310B>	BOPOMOFO LETTER N
5665	<l4>	<U310C>	BOPOMOFO LETTER L
5666	<g4>	<U310D>	BOPOMOFO LETTER G
5667	<k4>	<U310E>	BOPOMOFO LETTER K
5668	<h4>	<U310F>	BOPOMOFO LETTER H
5669	<j4>	<U3110>	BOPOMOFO LETTER J
5670	<q4>	<U3111>	BOPOMOFO LETTER Q
5671	<x4>	<U3112>	BOPOMOFO LETTER X
5672	<zh>	<U3113>	BOPOMOFO LETTER ZH
5673	<ch>	<U3114>	BOPOMOFO LETTER CH
5674	<sh>	<U3115>	BOPOMOFO LETTER SH
5675	<r4>	<U3116>	BOPOMOFO LETTER R
5676	<z4>	<U3117>	BOPOMOFO LETTER Z
5677	<c4>	<U3118>	BOPOMOFO LETTER C
5678	<s4>	<U3119>	BOPOMOFO LETTER S
5679	<a4>	<U311A>	BOPOMOFO LETTER A
5680	<o4>	<U311B>	BOPOMOFO LETTER O
5681	<e4>	<U311C>	BOPOMOFO LETTER E
5682	<eh4>	<U311D>	BOPOMOFO LETTER EH
5683	<ai>	<U311E>	BOPOMOFO LETTER AI
5684	<ei>	<U311F>	BOPOMOFO LETTER EI
5685	<au>	<U3120>	BOPOMOFO LETTER AU
5686	<ou>	<U3121>	BOPOMOFO LETTER OU
5687	<an>	<U3122>	BOPOMOFO LETTER AN
5688	<en>	<U3123>	BOPOMOFO LETTER EN
5689	<aN>	<U3124>	BOPOMOFO LETTER ANG
5690	<eN>	<U3125>	BOPOMOFO LETTER ENG
5691	<er>	<U3126>	BOPOMOFO LETTER ER
5692	<i4>	<U3127>	BOPOMOFO LETTER I
5693	<u4>	<U3128>	BOPOMOFO LETTER U
5694	<iu>	<U3129>	BOPOMOFO LETTER IU
5695	<v4>	<U312A>	BOPOMOFO LETTER V
5696	<nG>	<U312B>	BOPOMOFO LETTER NG
5697	<gn>	<U312C>	BOPOMOFO LETTER GN
5698	<(JU)>	<U321C>	PARENTHEZIZED HANGUL CIEUC U
5699	<1c>	<U3220>	PARENTHEZIZED IDEOGRAPH ONE
5700	<2c>	<U3221>	PARENTHEZIZED IDEOGRAPH TWO
5701	<3c>	<U3222>	PARENTHEZIZED IDEOGRAPH THREE
5702	<4c>	<U3223>	PARENTHEZIZED IDEOGRAPH FOUR
5703	<5c>	<U3224>	PARENTHEZIZED IDEOGRAPH FIVE
5704	<6c>	<U3225>	PARENTHEZIZED IDEOGRAPH SIX
5705	<7c>	<U3226>	PARENTHEZIZED IDEOGRAPH SEVEN
5706	<8c>	<U3227>	PARENTHEZIZED IDEOGRAPH EIGHT

5707	<9c>	<U3228>	PARENTHESIZED IDEOGRAPH NINE
5708	<10c>	<U3229>	PARENTHESIZED IDEOGRAPH TEN
5709	<KSC>	<U327F>	KOREAN STANDARD SYMBOL
5710	<am>	<U33C2>	SQUARE AM
5711	<pm>	<U33D8>	SQUARE PM
5712	<ff>	<UFB00>	LATIN SMALL LIGATURE FF
5713	<fi>	<UFB01>	LATIN SMALL LIGATURE FI
5714	<f1>	<UFB02>	LATIN SMALL LIGATURE FL
5715	<ffi>	<UFB03>	LATIN SMALL LIGATURE FFI
5716	<ff1>	<UFB04>	LATIN SMALL LIGATURE FFL
5717	<St>	<UFB05>	LATIN SMALL LIGATURE LONG S T
5718	<st>	<UFB06>	LATIN SMALL LIGATURE ST
5719	<3+;>	<UFE7D>	ARABIC SHADDA MEDIAL FORM
5720	<aM.>	<UFE82>	ARABIC LETTER ALEF WITH MADDA ABOVE FINAL FORM
5721	<aH.>	<UFE84>	ARABIC LETTER ALEF WITH HAMZA ABOVE FINAL FORM
5722	<ah.>	<UFE88>	ARABIC LETTER ALEF WITH HAMZA BELOW FINAL FORM
5723	<a+->	<UFE8D>	ARABIC LETTER ALEF ISOLATED FORM
5724	<a+.>	<UFE8E>	ARABIC LETTER ALEF FINAL FORM
5725	<b+->	<UFE8F>	ARABIC LETTER BEH ISOLATED FORM
5726	<b+.>	<UFE90>	ARABIC LETTER BEH FINAL FORM
5727	<b+,>	<UFE91>	ARABIC LETTER BEH INITIAL FORM
5728	<b+;>	<UFE92>	ARABIC LETTER BEH MEDIAL FORM
5729	<tm->	<UFE93>	ARABIC LETTER TEH MARBUTA ISOLATED FORM
5730	<tm.>	<UFE94>	ARABIC LETTER TEH MARBUTA FINAL FORM
5731	<t+->	<UFE95>	ARABIC LETTER TEH ISOLATED FORM
5732	<t+.>	<UFE96>	ARABIC LETTER TEH FINAL FORM
5733	<t+,>	<UFE97>	ARABIC LETTER TEH INITIAL FORM
5734	<t+;>	<UFE98>	ARABIC LETTER TEH MEDIAL FORM
5735	<tk->	<UFE99>	ARABIC LETTER THEH ISOLATED FORM
5736	<tk.>	<UFE9A>	ARABIC LETTER THEH FINAL FORM
5737	<tk,>	<UFE9B>	ARABIC LETTER THEH INITIAL FORM
5738	<tk;>	<UFE9C>	ARABIC LETTER THEH MEDIAL FORM
5739	<g+->	<UFE9D>	ARABIC LETTER JEEM ISOLATED FORM
5740	<g+.>	<UFE9E>	ARABIC LETTER JEEM FINAL FORM
5741	<g+,>	<UFE9F>	ARABIC LETTER JEEM INITIAL FORM
5742	<g+;>	<UFEA0>	ARABIC LETTER JEEM MEDIAL FORM
5743	<hk->	<UFEA1>	ARABIC LETTER HAH ISOLATED FORM
5744	<hk.>	<UFEA2>	ARABIC LETTER HAH FINAL FORM
5745	<hk,>	<UFEA3>	ARABIC LETTER HAH INITIAL FORM
5746	<hk;>	<UFEA4>	ARABIC LETTER HAH MEDIAL FORM
5747	<x+->	<UFEA5>	ARABIC LETTER KHAH ISOLATED FORM
5748	<x+.>	<UFEA6>	ARABIC LETTER KHAH FINAL FORM
5749	<x+,>	<UFEA7>	ARABIC LETTER KHAH INITIAL FORM
5750	<x+;>	<UFEA8>	ARABIC LETTER KHAH MEDIAL FORM
5751	<d+->	<UFEA9>	ARABIC LETTER DAL ISOLATED FORM
5752	<d+.>	<UFEAA>	ARABIC LETTER DAL FINAL FORM
5753	<dk->	<UFEAB>	ARABIC LETTER THAL ISOLATED FORM
5754	<dk.>	<UFEAC>	ARABIC LETTER THAL FINAL FORM
5755	<r+->	<UFEAD>	ARABIC LETTER REH ISOLATED FORM
5756	<r+.>	<UFEAE>	ARABIC LETTER REH FINAL FORM
5757	<z+->	<UFEAF>	ARABIC LETTER ZAIN ISOLATED FORM
5758	<z+.>	<UFEBO>	ARABIC LETTER ZAIN FINAL FORM
5759	<s+->	<UFEB1>	ARABIC LETTER SEEN ISOLATED FORM
5760	<s+.>	<UFEB2>	ARABIC LETTER SEEN FINAL FORM
5761	<s+,>	<UFEB3>	ARABIC LETTER SEEN INITIAL FORM
5762	<s+;>	<UFEB4>	ARABIC LETTER SEEN MEDIAL FORM
5763	<sn->	<UFEB5>	ARABIC LETTER SHEEN ISOLATED FORM
5764	<sn.>	<UFEB6>	ARABIC LETTER SHEEN FINAL FORM
5765	<sn,>	<UFEB7>	ARABIC LETTER SHEEN INITIAL FORM
5766	<sn/>	<UFEB8>	ARABIC LETTER SHEEN MEDIAL FORM
5767	<c+->	<UFEB9>	ARABIC LETTER SAD ISOLATED FORM
5768	<c+.>	<UFEBA>	ARABIC LETTER SAD FINAL FORM
5769	<c+,>	<UFEBB>	ARABIC LETTER SAD INITIAL FORM
5770	<c+;>	<UFEBC>	ARABIC LETTER SAD MEDIAL FORM
5771	<dd->	<UFEBD>	ARABIC LETTER DAD ISOLATED FORM
5772	<dd.>	<UFEBE>	ARABIC LETTER DAD FINAL FORM
5773	<dd,>	<UFEBF>	ARABIC LETTER DAD INITIAL FORM
5774	<dd;>	<UFEC0>	ARABIC LETTER DAD MEDIAL FORM
5775	<tj->	<UFEC1>	ARABIC LETTER TAH ISOLATED FORM
5776	<tj.->	<UFEC2>	ARABIC LETTER TAH FINAL FORM
5777	<tj,>	<UFEC3>	ARABIC LETTER TAH INITIAL FORM
5778	<tj;>	<UFEC4>	ARABIC LETTER TAH MEDIAL FORM
5779	<zH->	<UFEC5>	ARABIC LETTER ZAH ISOLATED FORM
5780	<zH.>	<UFEC6>	ARABIC LETTER ZAH FINAL FORM
5781	<zH,>	<UFEC7>	ARABIC LETTER ZAH INITIAL FORM
5782	<zH;>	<UFEC8>	ARABIC LETTER ZAH MEDIAL FORM
5783	<e+->	<UFEC9>	ARABIC LETTER AIN ISOLATED FORM
5784	<e+.>	<UFECA>	ARABIC LETTER AIN FINAL FORM
5785	<e+,>	<UFECB>	ARABIC LETTER AIN INITIAL FORM
5786	<e+;>	<UFECC>	ARABIC LETTER AIN MEDIAL FORM
5787	<i+->	<UFECD>	ARABIC LETTER GHAIN ISOLATED FORM
5788	<i+.>	<UFECE>	ARABIC LETTER GHAIN FINAL FORM
5789	<i+,>	<UFECF>	ARABIC LETTER GHAIN INITIAL FORM
5790	<i+;>	<UFED0>	ARABIC LETTER GHAIN MEDIAL FORM
5791	<f+->	<UFED1>	ARABIC LETTER FEH ISOLATED FORM
5792	<f+.>	<UFED2>	ARABIC LETTER FEH FINAL FORM
5793	<f+,>	<UFED3>	ARABIC LETTER FEH INITIAL FORM
5794	<f+;>	<UFED4>	ARABIC LETTER FEH MEDIAL FORM

5795	<q+->	<UFED5>	ARABIC LETTER QAF ISOLATED FORM
5796	<q+,>	<UFED6>	ARABIC LETTER QAF FINAL FORM
5797	<q+,>	<UFED7>	ARABIC LETTER QAF INITIAL FORM
5798	<q+;>	<UFED8>	ARABIC LETTER QAF MEDIAL FORM
5799	<k+->	<UFED9>	ARABIC LETTER KAF ISOLATED FORM
5800	<k+,>	<UFEDA>	ARABIC LETTER KAF FINAL FORM
5801	<k+,>	<UFEDB>	ARABIC LETTER KAF INITIAL FORM
5802	<k+;>	<UFEDC>	ARABIC LETTER KAF MEDIAL FORM
5803	<l+->	<UFEDD>	ARABIC LETTER LAM ISOLATED FORM
5804	<l+,>	<UFEDF>	ARABIC LETTER LAM FINAL FORM
5805	<l+,>	<UFEDG>	ARABIC LETTER LAM INITIAL FORM
5806	<l+;>	<UFEE0>	ARABIC LETTER LAM MEDIAL FORM
5807	<m+->	<UFEE1>	ARABIC LETTER MEEM ISOLATED FORM
5808	<m+,>	<UFEE2>	ARABIC LETTER MEEM FINAL FORM
5809	<m+,>	<UFEE3>	ARABIC LETTER MEEM INITIAL FORM
5810	<m+;>	<UFEE4>	ARABIC LETTER MEEM MEDIAL FORM
5811	<n+->	<UFEE5>	ARABIC LETTER NOON ISOLATED FORM
5812	<n+,>	<UFEE6>	ARABIC LETTER NOON FINAL FORM
5813	<n+,>	<UFEE7>	ARABIC LETTER NOON INITIAL FORM
5814	<n+;>	<UFEE8>	ARABIC LETTER NOON MEDIAL FORM
5815	<h+->	<UFEE9>	ARABIC LETTER HEH ISOLATED FORM
5816	<h+,>	<UFEEA>	ARABIC LETTER HEH FINAL FORM
5817	<h+,>	<UFEFB>	ARABIC LETTER HEH INITIAL FORM
5818	<h+;>	<UFECC>	ARABIC LETTER HEH MEDIAL FORM
5819	<w+->	<UFEED>	ARABIC LETTER WAW ISOLATED FORM
5820	<w+,>	<UFEFF>	ARABIC LETTER WAW FINAL FORM
5821	<j+->	<UFEFF>	ARABIC LETTER ALEF MAKSURA ISOLATED FORM
5822	<j+,>	<UFEF0>	ARABIC LETTER ALEF MAKSURA FINAL FORM
5823	<y+->	<UFEF1>	ARABIC LETTER YEH ISOLATED FORM
5824	<y+,>	<UFEF2>	ARABIC LETTER YEH FINAL FORM
5825	<y+,>	<UFEF3>	ARABIC LETTER YEH INITIAL FORM
5826	<y+;>	<UFEF4>	ARABIC LETTER YEH MEDIAL FORM
5827	<LM->	<UFEF5>	ARABIC LIGATURE LAM WITH ALEF WITH MADDA ABOVE ISOLATED FORM
5828	<LM,>	<UFEF6>	ARABIC LIGATURE LAM WITH ALEF WITH MADDA ABOVE FINAL FORM
5829	<lh->	<UFEF7>	ARABIC LIGATURE LAM WITH ALEF WITH HAMZA ABOVE ISOLATED FORM
5830	<lh,>	<UFEF8>	ARABIC LIGATURE LAM WITH ALEF WITH HAMZA ABOVE FINAL FORM
5831	<lh->	<UFEF9>	ARABIC LIGATURE LAM WITH ALEF WITH HAMZA BELOW ISOLATED FORM
5832	<lh,>	<UFEFA>	ARABIC LIGATURE LAM WITH ALEF WITH HAMZA BELOW FINAL FORM
5833	<la->	<UFEFB>	ARABIC LIGATURE LAM WITH ALEF ISOLATED FORM
5834	<la,>	<UFEFC>	ARABIC LIGATURE LAM WITH ALEF FINAL FORM
5835	<H->	<U0023>	NUMBER SIGN
5836	<!S>	<U0024>	DOLLAR SIGN
5837	<@>	<U0040>	COMMERCIAL AT
5838	<Oa>	<U0040>	COMMERCIAL AT
5839	<!C>	<U00A2>	CENT SIGN
5840	<L->	<U00A3>	POUND SIGN
5841	<Xo>	<U00A4>	CURRENCY SIGN
5842	<Y->	<U00A5>	YEN SIGN
5843	<!B>	<U00A6>	BROKEN BAR
5844	<So>	<U00A7>	SECTION SIGN
5845	<7!>	<U00AC>	NOT SIGN
5846	<9I>	<U00B6>	PILCROW SIGN
5847	<_-->	<U2500>	BOX DRAWINGS LIGHT HORIZONTAL
5848	<_=>	<U2501>	BOX DRAWINGS HEAVY HORIZONTAL
5849	<_!>	<U2502>	BOX DRAWINGS LIGHT VERTICAL
5850	<_V/>>	<U250C>	BOX DRAWINGS LIGHT DOWN AND RIGHT
5851	<_V<w>	<U2510>	BOX DRAWINGS LIGHT DOWN AND LEFT
5852	<_A/>>	<U2514>	BOX DRAWINGS LIGHT UP AND RIGHT
5853	<_A<>	<U2518>	BOX DRAWINGS LIGHT UP AND LEFT
5854	<_!/>>	<U251C>	BOX DRAWINGS LIGHT VERTICAL AND RIGHT
5855	<_!<>	<U2524>	BOX DRAWINGS LIGHT VERTICAL AND LEFT
5856	<_V->	<U252C>	BOX DRAWINGS LIGHT DOWN AND HORIZONTAL
5857	<_A->	<U2534>	BOX DRAWINGS LIGHT UP AND HORIZONTAL
5858	<_!->	<U253C>	BOX DRAWINGS LIGHT VERTICAL AND HORIZONTAL
5859	<_/->/>	<U2571>	BOX DRAWINGS LIGHT DIAGONAL UPPER RIGHT TO LOWER LEFT
5860	<_-<>	<U2572>	BOX DRAWINGS LIGHT DIAGONAL UPPER LEFT TO LOWER RIGHT
5861	<_-./>/>	<U25E2>	BLACK LOWER RIGHT TRIANGLE
5862	<_-.<>	<U25E3>	BLACK LOWER LEFT TRIANGLE
5863	<_d!>	<U266A>	EIGHTH NOTE
5864			

## 7 CONFORMANCE

### 7.1 FDCC-set

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

### 7.2 FDCC-set category

Conformance can be claimed for a category description against each of the clauses 4.3

5876 thru 4.12, and then the requirements of clause 4.1 are also met, and a  
5877 LC\_IDENTIFICATION category as described in clause 4.2 is specified.  
5878

5879 **7.3 Charmap**  
5880

5881 A charmap description is conforming to this Technical Report if it meets the requirements  
5882 in clause 5.  
5883

5884 **7.4 Repertoiremap**  
5885

5886 A repertoiremap description is conforming to this Technical Report if it meets the  
5887 requirements in clause 6.  
5888

5889  
5890  
5891  
5892       **Annex A**  
5893           (informative)

5894       **Differences from the ISO/IEC 9945-2 standard**  
5895  
5896  
5897

5898       This Technical Report originated from the locale and charmap specifications in the  
5899       ISO/IEC 9945-2 POSIX shell and utilities standard, and it intends to be backwards  
5900       compatible, so that what is conformant to that standard should also be conformant to this  
5901       Technical Report.

5902       A number of enhancements have been made and a number of restrictions have been lifted  
5903       in comparison to the POSIX standard:

5904       **A.1 Restrictions removed**  
5905  
5906

5907       1. Dependence on specific meaning of the character NUL as termination of a string (from  
5908       the C standard) has been removed, to cater for other programming languages than C.

5909       **A.2 Enhancements**  
5910  
5911

5912       1. A description of a "repertoireemap" definition was added to facilitate descriptions of  
5913       FDCC-sets without charmaps, and also to provide binding from a FDCC-set using one set  
5914       of character names to charmaps using another naming set.

5915       2. The specific POSIX locale has been replaced with the "i18n" FDCC-set, defined on the  
5916       repertoire on ISO/IEC 10646.

5917       3. Transliteration support has been added in the LC\_CTYPE category.

5918       4. Terminology has been aligned with ISO/IEC TR 11017, especially the POSIX term  
5919       "locale" has been changed to "FDCC-set".

5920       5. A date escape format "%F" has been added for ISO 8601 dates, and another date escape  
5921       format "%f" has been added for weekday number with Monday being the first day of the  
5922       week.

5923       6. Added to LC\_MONETARY to accommodate differences between local and international  
5924       formats:

5925           int\_p\_cs\_precedes  
5926           int\_p\_sep\_by\_space  
5927           int\_n\_cs\_precedes  
5928           int\_n\_sep\_by\_space

5929       7. Section symbols have been added via the "section-symbol" keyword in the  
5930       LC\_COLLATE category.

5931       8. The "order\_start" keyword has got an optional "section-symbol" identifier

5932       9. The keywords "reorder-section-after" and "reorder-section\_end" have been introduced to  
5933       reorder sections.

5934       10. Symbolic ellipses (both decimal and hexadecimal) has been introduced as a notation.

- 5941 11. The "print" CTYYPE class includes automatically all "graph" characters.
- 5942
- 5943 12. The <Uxxxx> and <Uxxxxxxxx> notations have been introduced as predefined
- 5944 symbolic character names, together with a number of symbolic character names derived
- 5945 from POSIX and the Internet.
- 5946
- 5947 13. New categories LC\_IDENTIFICATION, LC\_XLITERATE, LC\_NAME,
- 5948 LC\_ADDRESS, and LC\_TELEPHONE, have been introduced.
- 5949
- 5950 14. The LC\_CTYPE has got support for new classes, via the new keywords class and
- 5951 map, which corresponds to the C standard library functions iswctype() and towctrans()
- 5952 respectively.
- 5953
- 5954 15. The "digit" keyword now supports digits for multiple scripts.
- 5955
- 5956 16. The LC\_MONETARY category provides support for multiple currencies, such as the
- 5957 native currency and the Euro in some European countries.
- 5958
- 5959 17. The LC\_TIME has got a number of enhancements to cater for alternate calendars, and
- 5960 timezone information may be given.
- 5961
- 5962 18. The charmap specification has been enhanced to support ISO 2022.

5963                   **Annex B**  
5964                   (informative)

5965                   **Rationale**

5966                   **B.1 FDCC-set Rationale**

5967                   The description of FDCC-sets is based on work performed in the UniForum Technical  
5968                   Committee Subcommittee on Internationalisation and POSIX. Wherever appropriate,  
5969                   keywords were taken from the C Standard or the ISO/IEC 9945-2:1993 POSIX standard.  
5970                   The C and POSIX term "locale" has been changed into the term "FDCC-set" from  
5971                   ISO/IEC TR 11017 to align with that specification.

5972                   The POSIX utility "localedef" compiles locale sources into object files. The "object"  
5973                   definitions need not be portable, as long as "source" definitions are. Strictly speaking,  
5974                   "source" definitions are portable only between applications using the same character set(s).  
5975                   Such "source" definitions can, if they use symbolic names only, easily be ported between  
5976                   systems using different code sets as long as the characters in the portable character set  
5977                   (ISO 646) have common values between the code sets; this is frequently the case in  
5978                   historical applications. Of course, this requires that the symbolic names used for characters  
5979                   outside the portable character set are identical between character sets.

5980                   To avoid confusion between an octal constant and a backreference, the octal, hexadecimal,  
5981                   and decimal constants must contain at least two digits. As single-digit constants are  
5982                   relatively rare, this should not impose any significant hardship. Each of the constants  
5983                   includes "two or more" digits to account for systems in which the byte size is larger than  
5984                   eight bits. For example, an ISO/IEC 10646 system that has defined 16-bit bytes may  
5985                   require six octal, four hexadecimal, and five decimal digits, for some coded characters.

5986                   As an international (ISO/IEC) Technical Report this Technical Report should follow the  
5987                   ISO/IEC guidelines, including the ISO/IEC TR 10176. This TR has a rule that characters  
5988                   outside the invariant part of ISO/IEC 646 should not be used in portable specifications.  
5989                   The backslash and the number-sign character are not in the invariant part. As far as  
5990                   general usage of these symbols, they are covered by the "grandfather clause" specifying  
5991                   previous practise in international standards and in the industry such as in specifications  
5992                   from The Open Group, but for newly defined interfaces, ISO has requested that  
5993                   specifications provide alternate representations, and this Technical Report then follows  
5994                   POSIX for backward compatibility. Consequently, while the default escape character  
5995                   remains the backslash, and the default comment character is the number-sign, applications  
5996                   are required to recognize alternative representations, identified in the applicable source text  
5997                   via the "escape\_char" and "comment\_char" keywords.

6000                   **B.1.1 LC\_IDENTIFICATION Rationale.**

6001                   The LC\_IDENTIFICATION category gives meta-information on the FDCC-set, such as  
6002                   who created it, and what is the level of conformance for each of the FDCC sets.

6003                   **B.1.2 LC\_CTYPE Rationale**

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

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

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

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

### B.1.3 LC\_COLLATE Rationale.

The LC\_COLLATE category governs the collation order in the FDCC-set, and may thus be useful for the processing of the ISO/IEC 14651 string ordering and comparison standard, the C Standard strxfrm() and strcoll() functions, as well as a number of ISO/IEC 9945-2:1993 POSIX utilities.

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

- (1) Byte/machine code order. This is the historical collation order in the UNIX system and many proprietary operating systems. Collation is here done character by character, without any regard to context. The primary virtue is that it usually is quite fast, and also completely deterministic; it works well when the native machine collation sequence matches the user expectations.
- (2) Character order. On this level, collation is also done character by character, without regard to context. The order between characters is, however, not determined by the code values, but on the user's expectations of the correct order between characters. In addition, such a (simple) collation order can specify that certain characters collate equal (e.g., upper and lowercase letters).
- (3) String ordering. On this level, entire strings are compared based on relatively

straightforward rules. At this level, several "passes" may be required to determine the order between two strings. Characters may be ignored in some passes, but not in others; the strings may be compared in different directions; and simple string substitutions may be made before strings are compared. This level is best described as "dictionary" ordering; it is based on the spelling, not the pronunciation, or meaning, of the words.

- (4) Text search ordering. This is a further refinement of the previous level, best described as "telephone book ordering"; some common homonyms (words spelled differently but with same pronunciation) are collated together; numbers are collated as if spelled with words, and so on.
- (5) Semantic level ordering. Words and strings are collated based on their meaning; entire words (such as "the") are eliminated, the ordering is not deterministic. This may require special software, and is highly dependent on the intended use.

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

As a result, the goal for the collation support is at level 3. This also matches the requirements for the Canadian collation order standard, as well as other, known collation requirements for alphabetic scripts. It specifically rules out collation based on pronunciation rules, or based on semantic analysis of the text. The syntax for the LC\_COLLATE category source is the result of a cooperative effort between representatives for many countries and organizations working with international issues, such as UniForum, The Open Group, The Unicode Consortium Inc. and ISO, and it meets the requirements for level 3, and has been verified to produce the correct result with examples based on Canadian and Danish collation order.

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

- (1) Once special characters (punctuation) have been removed from original strings, the ordering is determined by scanning forward (left to right) [disregarding case and diacriticals].
- (2) In case of equivalence, special characters are once again removed from original strings and the ordering is determined scanning backward (starting from the rightmost character of the string and back), character by character, (disregarding case but considering diacriticals).
- (3) In case of repeated equivalence, special characters are removed again from original strings and the ordering is determined scanning forward, character by character, (considering both case and diacriticals).
- (4) If there is still an ordering equivalence after rules (1) through (3) have been applied, then only special characters and the position they occupy in the string are considered to determine ordering. The string that has a special character in the lowest position comes first. If two strings have a special character in the

same position, the character [with the lowest collation value] comes first. In case of equality, the other special characters are considered until there is a difference or all special characters have been exhausted.

It is estimated that the Technical Report covers the mechanisms to specify data to cover the requirements for all European languages, and Cyrillic and Middle Eastern scripts.

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

The character (and collating element) order is defined by the order in which characters and elements are specified between the `order_start` and `order_end` keywords. This character order is used in range expressions in regular expressions. Weights assigned to the characters and elements define the collation sequence; in the absence of weights, the character order is also the collation sequence.

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

This Technical Report adds a number of facilities over the ISO/IEC 9945:1993 POSIX standard, especially in the support for the ISO/IEC 10646 UCS character set. These extended facilities are in alignment with the ISO/IEC 14651 sorting standard. In addition to the facilities provided in ISO/IEC 14651, this specification contains mechanisms to put data into a FDCC-set environment, and has added facilities to sort sections differently, has facilities to reuse FDCC-sets in different notations via the "equivalence-symbol" keyword and tables.

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

Much work has been done on FDCC-sets, making them quite general. The ISO/IEC 9945-2:1993 POSIX standard introduced a "copy" command for all categories of the POSIX locale. This is useful for many purposes and it ensures that two FDCC-sets are equivalent for this category. A further step in building on previous FDCC-set work is defined in this Technical Report.

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

6171 the Latin collating sequence is the same, but a few characters are different.

6172  
6173 This Technical Report defines a FDCC-set defined on the character repertoire of the  
6174 ISO/IEC 10646 standard, in a character set independent way. The intention is that some of  
6175 the information from this FDCC-set will be acceptable in many cultures, and that it can  
6176 serve as the basis for modifications in other cultures, to obtain a culturally acceptable  
6177 specification. Using the "reorder-after" construct will also help improve the overview of  
6178 what the changes really are for implementers and other users.

6179  
6180 An example of the use of the "reorder-after" construct is the following. A default  
6181 international ordering for the Latin alphabet may be adequate for Danish, with the  
6182 exception of the collation rules for the letters Ü, ü, Æ, æ, Ä, ä, Ø, ø, Ö, ö, Å and å. By  
6183 applying the "reorder-after" construct, the Danish specification can be made more easily  
6184 by copying and reordering the existing international specification, rather than specifying  
6185 collation parameters for all Latin letters (with or without diacritics). There is no obligation  
6186 for Denmark to take this approach, but the "reorder-after" construct provides the  
6187 mechanism for doing so if it is deemed desirable.

6188

6189

### 6190 B.1.3.2 awk script for "reorder-after" construct

6191

6192 A script has been written in the "awk" language defined in the POSIX standard ISO/IEC  
 6193 9945-2 to implement the "reorder-after" construct. It functions as follows: It reads all of  
 6194 the FDCC-set and if in the LC\_COLLATE category, it processes the line, else it just  
 6195 outputs the line. For the LC\_COLLATE category it reads the lines and puts it into a  
 6196 double linked list of strings identified by a line number; at the end of the LC\_COLLATE  
 6197 category all the lines are output. If the line is a "copy" keyword and it reads the file  
 6198 referenced, extracting the LC\_COLLATE section of the file in to the list of strings. If the  
 6199 line is a "reorder-after" keyword, it sets a pointer to be the line number of the symbol to  
 6200 of the "reorder-after" keyword. If the line is part of the "reorder-after" specification, it is  
 6201 entered into the double linked list at this point, and the previous entry in the double linked  
 6202 list for the <collation-element> is removed from the list. A "reorder-end" keyword  
 6203 terminates the reordering.

6204

```

6205 BEGIN { comment = "%"; back[0]= follow[0] = 0; }
6206 /LC_COLLATE/ { coll=1 }
6207 /END LC_COLLATE/ { coll=0; for (lnr= 1; lnr; lnr= follow[lnr]) print c-
6208 ont[lnr] }
6209
6210 { if (coll == 0) print $0 ;
6211   else { if ($1 == "copy") {
6212     file = $2
6213     while (getline < file )
6214       if ( $1 == "LC_COLLATE" ) copy_lc = 1
6215       else if ( $1 == "END" && $2 == "LC_COLLATE" ) copy_lc = 0
6216       else if (copy_lc) {
6217         lnr++
6218         follow[lnr-1] = lnr; back [ lnr ] = lnr-1
6219         cont[lnr] = $0; symb[ $1 ] = lnr
6220       }
6221       close (file )
6222     }
6223   else if ($1 == "reorder-after") { ra=1 ; after = symb [ $2 ] }
6224   else if ($1 == "reorder-end") ra = 0
6225   else {
6226     lnr++
6227     if (ra) follow [ lnr ] = follow [ after ]
6228     if (ra) back [ follow [ after ] ] = lnr
6229     follow[after] = lnr; back [ lnr ] = after
6230     cont[lnr] = $0
6231     if ( ra && $1 != comment && $1 != "" ) {
6232       old = symb [ $1 ];
6233       follow [ back [ old ] ] = follow [ old ];
6234       back [ follow [ old ] ] = back [ old ];
6235       symb[ $1 ] = lnr;
6236     }
6237     after = lnr
6238   }
6239 }
6240 }
```

6241

```

6242    B.1.3.3 Sample FDCC-set specification for Danish
6243
6244    escape_char /
6245    comment_char %
6246    repertoiremap "i18nrep"
6247    charset "ISO_8859-1:1987"
6248    % Distribution and use is free, also
6249    % for commercial purposes.
6250
6251    LC_VERSION
6252    title      "Danish language FDCC-set for Denmark"
6253    source     "Danish Standards Association"
6254    address   "Kollegievej 6, DK-2920 Charlottenlund, Danmark"
6255    contact   "Keld Simonsen"
6256    email     "Keld.Simonsen@dkuug.dk"
6257    tel       "+45 - 3996-6101"
6258    fax       "+45 - 3996-6202"
6259    language  "da"
6260    territory "DK"
6261    revision  "4.2"
6262    date     "1997-12-22"
6263
6264    category  i18n:2000;LC_IDENTIFICATION
6265    category  i18n:2000;LC_CTYPE
6266    category  i18n:2000;LC_COLLATE
6267    category  i18n:2000;LC_TIME
6268    category  posix:1993;LC_NUMERIC
6269    category  i18n:2000;LC_MONETARY
6270    category  posix:1993;LC_MESSAGES
6271    category  i18n:2000;LC_XLITERATE
6272    category  i18n:2000;LC_NAME
6273    category  i18n:2000;LC_ADDRESS
6274    category  i18n:2000;LC_TELEPHONE
6275
6276    END LC_VERSION
6277
6278    LC_CTYPE
6279    copy "i18n"
6280    END LC_CTYPE
6281
6282    LC_COLLATE
6283    % The ordering algorithm is in accordance
6284    % with Danish Standard DS 377 (1980)
6285    % and the Danish Orthography Dictionary
6286    % (Retskrivningsordbogen, 2. udgave, 1996).
6287    % It is also in accordance with
6288    % Greenlandic orthography.
6289
6290    collating-element <A-A> from "<A><A>"           "
6291    collating-element <A-a> from "<A><a>"           "
6292    collating-element <a-A> from "<a><A>"           "
6293    collating-element <a-a> from "<a><a>"           "
6294    collating-symbol <SPECIAL>
6295    copy i18n
6296    reorder-after <CAPITAL>
6297    <CAPITAL>
6298    <CAPITAL-SMALL>
6299    <SMALL-CAPITAL>
6300    <SMALL>
6301    reorder-after <q8>
6302    <kk>      <Q>;<SPECIAL>;<SMALL>;IGNORE
6303    reorder-after <t8>
6304    <TH>      "<T><H>" ; "<TH><TH>" ; "<CAPITAL><CAPITAL>" ; IGNORE
6305    <th>      "<T><H>" ; "<TH><TH>" ; "<SMALL><SMALL>" ; IGNORE
6306    reorder-after <y8>
6307    % <U:> and <U"> are treated as <Y> in Danish
6308    <U:>      <Y>;<U:>;<CAPITAL>;IGNORE
6309    <u:>      <Y>;<U:>;<SMALL>;IGNORE
6310    <U">      <Y>;<U">;<CAPITAL>;IGNORE
6311    <u">      <Y>;<U">;<SMALL>;IGNORE

```

```

6312 reorder-after <z8>
6313 % <AE> is a separate letter in Danish
6314 <AE>     <AE>;<NONE>;<CAPITAL>;IGNORE
6315 <ae>      <AE>;<NONE>;<SMALL>;IGNORE
6316 <AE'>    <AE>;<ACUTE>;<CAPITAL>;IGNORE
6317 <ae'>    <AE>;<ACUTE>;<SMALL>;IGNORE
6318 <A3>      <AE>;<MACRON>;<CAPITAL>;IGNORE
6319 <a3>      <AE>;<MACRON>;<SMALL>;IGNORE
6320 <A:>      <AE>;<SPECIAL>;<CAPITAL>;IGNORE
6321 <a:>      <AE>;<SPECIAL>;<SMALL>;IGNORE
6322 % <O//> is a separate letter in Danish
6323 <O//>    <O//>;<NONE>;<CAPITAL>;IGNORE
6324 <o//>    <O//>;<NONE>;<SMALL>;IGNORE
6325 <O//>'    <O//>;<ACUTE>;<CAPITAL>;IGNORE
6326 <o//>'    <O//>;<ACUTE>;<SMALL>;IGNORE
6327 <O:>      <O//>;<DIAERESIS>;<CAPITAL>;IGNORE
6328 <o:>      <O//>;<DIAERESIS>;<SMALL>;IGNORE
6329 <O">      <O//>;<DOUBLE-ACUTE>;<CAPITAL>;IGNORE
6330 <o">      <O//>;<DOUBLE-ACUTE>;<SMALL>;IGNORE
6331 % <AA> is a separate letter in Danish
6332 <AA>      <AA>;<NONE>;<CAPITAL>;IGNORE
6333 <aa>      <AA>;<NONE>;<SMALL>;IGNORE
6334 <A-A>    <AA>;<A-A>;<CAPITAL>;IGNORE
6335 <A-a>    <AA>;<A-A>;<CAPITAL-SMALL>;IGNORE
6336 <a-A>    <AA>;<A-A>;<SMALL-CAPITAL>;IGNORE
6337 <a-a>    <AA>;<A-A>;<SMALL>;IGNORE
6338 <AA'>    <AA>;<AA'>;<CAPITAL>;IGNORE
6339 <aa'>    <AA>;<AA'>;<SMALL>;IGNORE
6340 reorder-end
6341 END LC_COLLATE
6342
6343 LC_MONETARY
6344 int_curr_symbol      "<D><K><K><SP> "
6345 currency_symbol       "<k><r> "
6346 mon_decimal_point     "<,> "
6347 mon_thousands_sep      "<. > "
6348 mon_grouping          "3;3"
6349 positive_sign         ""
6350 negative_sign         "<-> "
6351 int_frac_digits        2
6352 frac_digits            2
6353 p_cs_precedes          1
6354 p_sep_by_space          2
6355 n_cs_precedes          1
6356 n_sep_by_space          2
6357 p_sign_posn            4
6358 n_sign_posn            4
6359
6360 END LC_MONETARY
6361
6362 LC_NUMERIC
6363 decimal_point           "<, > "
6364 thousands_sep            "<. > "
6365 grouping                "3;3"
6366
6367 LC_TIME
6368 abday      "<m><a><n>" ; /
6369                 "<t><i><r>" ; "<o><n><s>" ; /
6370                 "<t><o><r>" ; "<f><r><e>" ; /
6371                 "<l><o//><r>" ; "<s><o/><n>"
6372 day        "<m><a><n><d><a><g>" ; /
6373                 "<t><i><r><s><d><a><g>" ; /
6374                 "<o><n><s><d><a><g>" ; /
6375                 "<t><o><r><s><d><a><g>" ; /
6376                 "<f><r><e><d><a><g>" ; /
6377                 "<l><o//><r><d><a><g>" ;
6378                 "<s><o//><n><d><a><g>" ;
6379 week       7;19971201;4
6380 abmon      "<j><a><n>" ; "<f><e><b>" ; /
6381                 "<m><a><r>" ; "<a><p><r>" ; /

```

```

6382
6383
6384
6385
6386 mon
6387
6388
6389
6390
6391
6392
6393
6394
6395
6396
6397
6398 d_t_fmt
6399 d_fmt
6400 atl_digits
6401
6402
6403
6404
6405
6406
6407 t_fmt
6408 am_pm
6409 t_fmt_ampm
6410 timezone
6411
6412 END LC_TIME
6413
6414 LC_MESSAGES
6415 yesexpr " <<(><1><J><j><Y><y><) />><. ><*> "
6416 noexpr " <<(><0><N><n><) />><. ><*> "
6417
6418 END LC_MESSAGES
6419
6420 LC_NAME
6421 name_fmt "<%><p><%><t><%><g><%><t><%><m><%><t><%><f> "
6422 name_gen ""
6423 name_mr "<h><r> "
6424 name_mrs "<f><r><u> "
6425 name_miss "<f><r><o/><k><e><n> "
6426 name_ms "<f><r> "
6427
6428 LC_ADDRESS
6429 country_name "<D><a><n><m><a><r><k> "
6430 country_post "<D><K> "
6431 lang_ab "<d><a> "
6432 lang_term "<d><a><n> "
6433 postal_fmt "<%><a><%><N><%><f><%><N><%><d><%><N><%><b><%><N><%>/
6434 <%><s><SP><%><h><SP><%><e><SP><%><r><%><N> /
6435 <%><C><-><%><z><SP><%><T><%><N><%><c><%><N> "
6436
6437 END LC_ADDRESS
6438
6439 LC_TELEPHONE
6440 tel_int_fmt "<+><%><c><SP><%><a><SP><%><l> "
6441 tel_dom_fmt "<%><l> "
6442 int_select "<0><0> "
6443 int_prefix "<4><5> "
6444
6445 B.1.4 LC_MONETARY Rationale.
6446

```

6447 The currency symbol does not appear in LC\_MONETARY because it is not defined in the  
6448 C Standard's C locale. The C Standard limits the size of decimal points and thousands  
6449 delimiters to single-byte values. In FDCC-sets based on multibyte coded character sets this

cannot be enforced, obviously; this Technical Report does not prohibit such characters, but makes the behaviour unspecified (in the text "In contexts where other standards . . .").

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

The FDCC-set definition is an extension of the C Standard `localeconv()` specification. In particular, rules on how currency\_symbol is treated are extended to also cover `int_curr_symbol`, and `p_sep_by_space` and `n_sep_by_space` have been augmented with the value 2, which places a space between the sign and the symbol (if they are adjacent; otherwise it should be treated as a 0). The following table shows the result of various combinations:

		<code>p_sep_by_space</code>	2	1	0
6468	p_cs_precedes = 1	p_sign_posn = 0	(\$ 1.25)	(\$ 1.25)	(\$1.25)
6469		p_sign_posn = 1	+ \$1.25	+\$ 1.25	+\$1.25
6470		p_sign_posn = 2	\$1.25 +	\$ 1.25+	\$1.25+
6471		p_sign_posn = 3	+ \$1.25	+\$ 1.25	+\$1.25
6472		p_sign_posn = 4	\$ +1.25	\$+ 1.25	\$+1.25
6473					
6474	p_cs_precedes = 0	p_sign_posn = 0	(1.25 \$)	(1.25 \$)	(1.25\$)
6475		p_sign_posn = 1	+1.25 \$	+1.25 \$	+1.25\$
6476		p_sign_posn = 2	1.25\$ +	1.25 \$+	1.25\$+
6477		p_sign_posn = 3	1.25+ \$	1.25 +\$	1.25+\$
6478		p_sign_posn = 4	1.25\$ +	1.25 \$+	1.25\$+
6479					
6480					

The following is an example of the interpretation of the mon\_grouping keyword. Assuming that the value to be formatted is 123456789 and the mon\_thousands\_sep is "", then the following table shows the result. The third column shows the equivalent C Standard string that would be used to accommodate this grouping. It is the responsibility of the utility to perform mappings of the formats in this clause to those used by language bindings such as the C Standard .

Mon_grouping	Formatted Value	C String
3;-1	123456'789	"\3\177"
3	123'456'789	"\3"
3;2;-1	1234'56'789	"\3\2\177"
3;2	12'34'56'789	"\3\2"
-1	123456789	"177"

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

The multiple currency support is specified such that a FDCC-set can be used without change during the transition period in a static environment. For example in the case of the Euro currency as being employed in a number of European countries, there is no need to change the FDCC-set when shifting from one currency to two concurrent currencies; and

6502 there is no need to change FDCC-set, when changing to the Euro as the only currency.  
 6503 Also the same application call can be made to be valid for countries with a single  
 6504 currency and countries with dual currencies. The specifications can also be used without  
 6505 change of the FDCC-set on an installation, when converting from one national currency to  
 6506 another, for example when removing some zeroes to form a new currency.

6507  
 6508 The following example illustrates the support for multiple currencies; the example is for  
 6509 the Euro in Germany:

```

6510 LC_MONETARY
6511 valid_from      " " ;           "19990101"
6512 valid_to        "20020630" ;    ""
6513 conversion_rate 1 ;           "195/100
6514 int_curr_symbol "<D><E><M><SP>" ; "<E><U><R><SP>"
6515 currency_symbol  "<D><M>" ;      "<E><U><R>""
6516 mon_decimal_point "<,>" ;
6517 mon_thousands_sep  "<.>" ;
6518 mon_grouping     3;3 ;
6519 positive_sign    " "
6520 negative_sign   "<->" ;
6521 int_frac_digits 2 ;           2
6522 frac_digits      2 ;           2
6523 p_cs_precedes   1 ;           1
6524 p_sep_by_space   2 ;           2
6525 n_cs_precedes   1 ;           1
6526 n_sep_by_space   2 ;           2
6527 p_sign_posn     4 ;           4
6528 n_sign_posn     4 ;           4
6529
6530
6531 END LC_MONETARY
6532

```

### 6533 B.1.5 LC\_NUMERIC Rationale.

6534 See the rationale for LC\_MONETARY (B1.3) for a description of the behaviour of  
 6535 grouping.

### 6536 B.1.6 LC\_TIME Rationale.

6537 The LC\_TIME descriptions of abday, day, and abmon imply a Gregorian style calendar  
 6538 (7-day weeks, 12-month years, leap years, etc.). Other calendars can be supported, for  
 6539 example calendars with a fixed week length.

6540 In some FDCC-sets the field descriptors for weekday and month names will be given with  
 6541 an initial small letter. Programs using these fields may need to adjust the capitalization if  
 6542 the output is going to be used at the beginning of a sentence.

6543 The field descriptors corresponding to the optional keywords consist of a modifier  
 6544 followed by a traditional field descriptor (for instance %Ex). If the optional keywords are  
 6545 not supported by the application or are unspecified for the current FDCC-set, these field  
 6546 descriptors are treated as the traditional field descriptor. For instance, assume the  
 6547 following keywords:

```

6548 alt_digits "0th";"1st";"2nd";"3rd";"4th";"5th";"6th";"7th";"8th";"9th";"10th"
6549 d_fmt "The %Od day of %B in %Y"
6550

```

6551 On 7/4/1776, the %x field descriptor would result in "The 4th day of July in 1776," while  
 6552 7/14/1789 would come out as "The 14 day of July in 1789." It can be noted that the above  
 6553 example is for illustrative purposes only; the %o modifier is primarily intended to provide

for Kanji or Hindi digits in date formats. While it is clear that an alternate year format is required, there is no consensus on the format or the requirements. As a result, while these keywords are reserved, the details are left unspecified. It is expected that National Standards Bodies will provide specifications.

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

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

### **B.1.8 LC\_XLITERATE Rationale.**

Transliteration is often language dependent, transliterating one specific language to another specific language. For example transliteration from Russian to English, and from Serbian to German would normally be quite different, although the same repertoire of characters would be transliterated. Even transliteration of two languages using the same script into one language (for example from Russian to Danish and from Serbian to Danish), or transliteration of the same language (for example Russian into English or German) may be different. The language to be transliterated to is identified with the FDCC-set, which may also be used to identify a specific language to be transliterated from. Transliteration may also be to a specific repertoire of characters, determined for example by limitations of displaying equipment, or what the user can intelligibly read. The capabilities here allows for multiple fallback, so that the specification can be valid for all target character repertoires, eliminating the need for specific data for each target repertoire.

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

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

The "profession" entry is intended for either the general profession of the person in question, or the job title, for use in letters or as part of the address on an envelope.

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

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

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

6612  
6613 The LC\_TELEPHONE category gives information to prepare a text for writing a telephone  
6614 number. The information is intended to be given to an API that has the various  
6615 information on a telephone number as parameters and yields a formatted string as the  
6616 return value. Both an international and a domestic formatting possibility is available.  
6617  
6618

## B.2 Character Set Rationale.

6621 This Technical Report poses no requirement that multiple character sets or code sets be  
6622 supported, leaving this as a marketing differentiation for implementors. Although multiple  
6623 charmaps are supported, it is the responsibility of the application to provide the file(s); if  
6624 only one is provided, only that one will be accessible.  
6625

6626 The character set description text provides the capability to describe character set attributes  
6627 (such as collation order or character classes) independent of character set encoding, and  
6628 using only the characters in the portable character set. This makes it possible to create  
6629 "generic" FDCC-set source texts for all code sets that share the portable character set  
6630 (such as the ISO/IEC 8859 family or IBM Extended ASCII).  
6631

6632 Applications are free to describe more than one code set in a character set description text.  
6633 For example, if an application defines ISO/IEC 8859-1 as the primary code set, and  
6634 ISO/IEC 8859-2 as an alternate set, with each character from the alternate code set  
6635 preceded in data by a shift code, a character set description text could contain a complete  
6636 description of the primary set and those characters from the secondary that are not  
6637 identical, the encoding of the latter including the shift code.  
6638

6639 Applications are free to choose their own symbolic names, as long as the names identified  
6640 by this Technical Report are also defined; this provides support for already existing  
6641 "character names".  
6642

6643 The charmap was introduced to resolve problems with the portability of, especially,  
6644 FDCC-set sources. While the portable character set (in Table 1) is a constant across all  
6645 FDCC-sets for a particular application, this is not true for the extended character set.  
6646 However, the particular coded character set used for an application does not necessarily  
6647 imply different characteristics or collation: on the contrary, these attributes should in many  
6648 cases be identical, regardless of codeset. The charmap provides the capability to define a  
6649 common FDCC-set definition for multiple codesets (the same FDCC-set source can be  
6650 used for codesets with different extended characters; the ability in the charmap to define  
6651 "empty" names allows for characters missing in certain codesets).  
6652

6653 In addition, some implementors have expressed an interest in using the charmap to define  
6654 certain other characteristics of codesets, such as the <mb\_cur\_max> value for the  
6655 particular codeset. (Note that <mb\_cur\_max> has to be equal to or lower than the C  
6656 Standard {MB\_LEN\_MAX}, which is the application limit). Such extensions are not  
6657 described here; but may be added in a later revision of this Technical Report.  
6658

6659 The <escape\_char> declaration was added at the request of the international community to  
6660 ease the creation of portable charmaps on terminals not implementing the default  
6661 backslash escape. (This approach was adopted because this is a new interface invented by  
6662 ISO/IEC 9945-2:1993 POSIX. Historical interfaces, such as the shell command language  
6663 and awk, have not been modified to accommodate this type of terminal.)

6664 The octal number notation was selected to match those of POSIX "awk" and "tr" utilities  
6665 and is consistent with that used by the POSIX localedef utility.

6666  
6667 The charmap capability implements a facility available at some X/Open compatible  
6668 applications. Its prime virtue is to support "generic" collation sequence source definitions.  
6669 An implementor or an applications developer can produce a template definition that can be  
6670 used to produce several codeset-dependent "compiled" FDCC-set definitions. The facility  
6671 also removes any dependency in many source definitions on characters outside the  
6672 character set defined in this clause.

6673  
6674 The charmap allows specification of more than one encoding of a character. This allows  
6675 for encodings that can encode items in more than one way. For example, an item can be  
6676 encoded once as a fully composed character and again as a base character plus combining  
6677 character. This would allow either representation to be recognized. As only the first  
6678 occurrence of the character may be output, this technique could be used to normalize a  
6679 character stream.

6680  
6681 The ISO 2022 support introduced gives the possibility to refer other definitions via  
6682 charmaps, so the full encoding does not have to be replicated. It supports shifting with G0,  
6683 G1, G2 and G3 sets, and also general shifting of coded character sets via escape  
6684 sequences.

6685  
6686  
6687 **B.3 Repertoiremap Rationale.**

6688  
6689 The repertoiremap was introduced to make FDCC-sets independent of the availability of  
6690 charmaps. With the repertoiremap it is possible to use a FDCC-set encoded with one set  
6691 of symbolic character names, together with charmaps with other symbolic character  
6692 naming schemes, provided there are repertoiremaps available for both naming schemes.

6693  
6694 Repertoiremaps are also useful to describe repertoires of characters, to be used for  
6695 example for transliteration.

6696

6697

6698

6699

6700

6701

6702

## C.1 BNF Syntax Rules

6703

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.

6706

Definitions between <angle brackets> make use of terms not defined in this BNF syntax, and assume general English usage.

6709

6710

Other conventions:

6711

\* means 0 or more repetitions of a token.

6712

+ means one or more repetitions of a token

6713

Brackets [ ] indicate optional occurrence of a token.

6714

Comments start with a % on a separate line.

6715

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

6718

## C.2 Grammar for FDCC-sets

6719

6720

```
% The following is the overall FDCC-set grammar
FDCC_set_definition      = [ global_statement* ] category+ ;
global_statement          = 'escape_char' SP char_symbol EOL
                           | 'comment_char' SP char_symbol EOL
                           | 'repertoiremap' SP quoted_string EOL
                           | 'charmap' SP quoted_string EOL ;
category                   = lc_identification | lc_ctype | lc_collate
                           | lc_monetary | lc_numeric | lc_time
                           | lc_messages | lc_xliterate | lc_telephone
                           | lc_name | lc_address ;

% The following is the LC_IDENTIFICATION category grammar
lc_ident                  = ident_head ident_keyword* ident_tail ;
                           | ident_head copy_FDCC_set ident_tail ;
ident_head                 = 'LC_IDENTIFICATION' EOL ;
ident_keyword              = ident_keyword_string SP quoted_string EOL ;
                           = 'title' | 'source' | 'address' | 'contact'
                           | 'email' | 'tel' | 'fax' | 'language'
                           | 'territory' | 'audience' | 'application'
                           | 'abbreviation' | 'revision' | 'date' ;
ident_tail                 = 'END' SP 'LC_IDENTIFICATION' EOL ;

% The following is the LC_CTYPE category grammar
lc_ctype                   = ctype_head ctype_keyword* ctype_tail ;
                           | ctype_head copy_FDCC_set ctype_tail ;
ctype_head                 = 'LC_CTYPE' EOL ;
ctype_keyword              = charclass_keyword SP charclass_list EOL
                           | charconv_keyword SP charconv_list EOL
                           | 'width' SP width_list EOL ;
                           = 'upper' | 'lower' | 'alpha' | 'digit' |
                           | 'alnum' | 'punct' | 'xdigit' | 'space' |
                           | 'print' | 'graph' | 'blank' | 'cntrl' |
                           | 'outdigit'
                           | 'class' charclass_name semicolon ;
                           = '"combining"' | '"combining_level3"'
                           | ''' identifier ''' ;
```

```

6758 charclass_list
6759 = charclass_list semicolon char_symbol
6760 | charclass_list semicolon ctype_abs_ellipsis
6761 semicolon char_symbol
6762 | charclass_list semicolon charsymbol
6763 ctype_symbolic_ellipses charsymbol
6764 | char_symbol ;
6765 width_list
6766 = charclass_list ':' number
6767 | width_list semicolon width_list ;
6768 = 'toupper' | 'tolower'
6769 | 'map' ''' identifier ''' semicolon ;
6770 = charconv_list semicolon charconv_entry
6771 | charconv_entry ;
6772 = (' char_symbol comma char_symbol ') ;
6773 = '...' | '....' | '..(2)...' ;
6774 = '....' ;
6775 = 'END' SP 'LC_TYPE' EOL ;

6776 % The following is the LC_COLLATE category grammar
6777 lc_collate
6778 collate_head
6779 collate_keywords
6780 opt_statement
6781
6782
6783
6784
6785
6786
6787 collelem_string
6788 order_statements
6789 order_start
6790
6791
6792 order_opts
6793 order_opt
6794 opt_word
6795 collation_order
6796 collation_statement
6797
6798 collsymbol_list
6799
6800 collsymbol_elemnt
6801
6802 collating_element
6803
6804 weight_list
6805 weight_symbol
6806
6807
6808
6809
6810 ellipses
6811 reorder_after
6812 reorder_end
6813 reorder_section_after
6814
6815 reorder_section_end
6816 order_end
6817 collate_tail
6818
6819 % The following is the LC_MESSAGES category grammar
6820 lc_messages
6821 = messages_head messages_keyword* messages_tail
6822 | messages_head copy_FDCC_set messages_tail ;
6823 = 'LC_MESSAGES' EOL ;
6824 = 'yesexpr' SP '"' extended_reg_expr '"' EOL
6825 | 'yesexpr' SP ''' extended_reg_expr ''' EOL ;
6826 = 'END' SP 'LC_MESSAGES' EOL ;
6827
6828 % The following is the LC_MONETARY category grammar

```

```

6828 lc_monetary
6829 monetary_head monetary_keyword*
6830 monetary_head copy_FDCC_set monetary_tail ;
6831 = 'LC_MONETARY' EOL ;
6832 mon_keyword_string SP quoted_string EOL
6833 | mon_keyword_strings SP mon_string_list EOL
6834 | mon_keyword_char SP mon_number_list EOL
6835 | mon_keyword_date SP mon_date_list EOL
6836 | 'conversion_rate' SP mon_conv_list EOL
6837 | 'mon_grouping' SP mon_group_list EOL ;
6838 = 'mon_decimal_point' | 'mon_thousands_sep'
6839 | 'positive_sign' | 'negative_sign' ;
6840 = 'int_curr_symbol' | 'currency_symbol' ;
6841 = 'int_frac_digits' | 'frac_digits'
6842 | 'p_cs_precedes' | 'p_sep_by_space'
6843 | 'n_cs_precedes' | 'n_sep_by_space'
6844 | 'int_p_cs_precedes' | 'int_p_sep_by_space'
6845 | 'int_n_cs_precedes' | 'int_n_sep_by_space'
6846 | 'p_sign_posn' | 'n_sign_posn'
6847 | 'int_p_sign_posn' | 'int_n_sign_posn' ;
6848 = 'valid_from' | 'valid_to' ;
6849 = mon_date | mon_date_list semicolon mon_date ;
6850 = ''' 8 * digit ''' ;
6851 = number | mon_group_list semicolon number ;
6852 = quoted_string [ semicolon quoted_string]* ;
6853 = mon_number | mon_number_list semicolon
6854 mon_number ;
6855 = number | -1 ;
6856 = mon_pair | mon_conv_list semicolon mon_pair ;
6857 = number spaces* '/' spcaes* number ;
6858 = 'END' SP 'LC_MONETARY' EOL ;

6859 % The following is the LC_NUMERIC category grammar
6860 lc_numeric
6861 numeric_head numeric_keyword*
6862 numeric_head copy_FDCC_set numeric_tail ;
6863 = 'LC_NUMERIC' EOL ;
6864 num_keyword_string SP quoted_string EOL
6865 | num_keyword_grouping SP num_group_list EOL ;
6866 = 'decimal_point' | 'thousands_sep' ;
6867 = 'grouping' ;
6868 = number
6869 | num_group_list semicolon number ;
6870 = 'END' SP 'LC_NUMERIC' EOL ;

6871 % The following is the LC_TIME category grammar
6872 lc_time
6873 time_head time_keyword*
6874 time_head copy_FDCC_set time_tail ;
6875 = 'LC_TIME' EOL ;
6876 time_keyword_name SP time_list EOL
6877 | time_keyword_fmt SP quoted_string EOL
6878 | time_keyword_opt SP time_list EOL
6879 | 'week' SP number semicolon mon_date semicolon
6880 number EOL
6881 | time_keyword_num SP number EOL
6882 | 'timezone' SP time_list EOL;
6883 = 'abday' | 'day' | 'abmon' | 'mon' | 'am_pm' ;
6884 = 'd_t_fmt' | 'd_fmt' | 't_fmt' | 't_fmt_ampm';
6885 = 'era' | 'era_year' | 'era_d_fmt' | 'alt_digits'
6886 | era_d_t_fmt | era_t_fmt ;
6887 = 'week' ;
6888 = 'first_weekday' | 'first_workday'
6889 | 'cal_direction' ;
6890 = time_list semicolon quoted_string
6891 | quoted_string ;
6892 = 'END' SP 'LC_TIME' EOL ;

6893 % The following is the LC_XLITERATE category grammar
6894 lc_xliterate
6895 = translit_head [translit_include]
6896 [default_missing] translit_statement*
6897 translit_tail | translit_head copy_FDCC_set
6898 translit_tail ;

```

```

6898 translit_head          = 'LC_XLITERATE' EOL ;
6899 translit_include        = 'include' SP FDCC_set_name semicolon
6900
6901 default_missing         = 'default_missing' SP quoted_string EOL ;
6902 translit_ignore          = 'translit_ignore' SP charclass_list EOL ;
6903 translit_statement        = char_or_string SP char_or_string [ semicolon
6904                               char_or_string ]* EOL ;
6905 translit_tail            = 'END' SP 'LC_XLITERATE' EOL ;
6906
6907 % The following is the LC_NAME category grammar
6908 lc_name                  = name_head name_keyword* name_tail
6909                               | name_head copy_FDCC_set name_tail ;
6910
6911 name_head                = 'LC_NAME' EOL ;
6912 name_keyword              = name_keyword_string SP quoted_string EOL ;
6913 name_keyword_string        = 'name_fmt' | 'name_gen' | 'name_mr'
6914                               | 'name_mrs' | 'name_ms' | 'name_miss'
6915                               | 'name_ms' ;
6916 name_tail                 = 'END' SP 'LC_NAME' EOL ;
6917
6918 % The following is the LC_ADDRESS category grammar
6919 lc_address                = address_head address_keyword* address_tail
6920                               | address_head copy_FDCC_set address_tail ;
6921 address_head               = 'LC_ADDRESS' EOL ;
6922 address_keyword            = address_keyword_string SP quoted_string EOL ;
6923 address_keyword_string      = 'postal_fmt' | 'country_name' |
6924                               | 'country_post' | 'lang_name' | 'lang_ab2' |
6925                               | 'lang_ab3_term' | 'lang_ab3_lib' ;
6926 address_tail               = 'END' SP 'LC_ADDRESS' EOL ;
6927
6928 % The following is the LC_TELEPHONE category grammar
6929 lc_tel                     = tel_head tel_keyword* tel_tail
6930                               | tel_head copy_FDCC_set tel_tail ;
6931 tel_head                   = 'LC_TELEPHONE' EOL ;
6932 tel_keyword                = tel_keyword_string SP quoted_string EOL ;
6933 tel_keyword_string          = 'tel_int_fmt' | 'tel_dom_fmt' | 'int_select'
6934                               | 'int_prefix' ;
6935 tel_tail                   = 'END' SP 'LC_TELEPHONE' EOL ;
6936
6937 % The following grammar rules are common to all categories
6938 char                        = <any character except those that makes an End
6939                               Of Line>
6940 graphic_char                = <any char except control_chars and space> ;
6941 space                         = ' ' | <TAB> ;
6942 SP                            = space+ ;
6943 EOL                           = end_of_line | comment_end_of_line ;
6944 end_of_line                   = <anything that makes an End Of Line (EOL) in
6945                               the operating system employed> ;
6946 comment_char                 = <defined by the 'comment_char' keyword> ;
6947 escape_char                  = <defined by the 'escape_char' keyword> ;
6948 charsymbol                   = simple_symbol | ucs_symbol ;
6949 collsymbol                   = simple_symbol ;
6950 collelement                  = simple_symbol ;
6951 sectionsymbol               = simple_symbol ;
6952 octdigit                     = '0' | '1' | '2' | '3' | '4' | '5' | '6' | '7' ;
6953 digit                         = '0' | '1' | '2' | '3' | '4' | '5' | '6' | '7' | '8' | '9' ;
6954 hex_upper                     = 'A' | 'B' | 'C' | 'D' | 'E' | 'F' | digit ;
6955 hexdigit                     = hex_upper | 'a' | 'b' | 'c' | 'd' | 'e' | 'f' | 'g' | 'h' | 'i' | 'j' | 'k'
6956                               | 'l' | 'm' | 'n' | 'o' | 'p' | 'q' | 'r' | 's' |
6957                               | 't' | 'u' | 'v' | 'w' | 'x' | 'y' | 'z' | 'A' | 'B' | 'C' | 'D'
6958                               | 'E' | 'F' | 'G' | 'H' | 'I' | 'J' | 'K' | 'L' | 'M' | 'N' | 'O'
6959                               | 'P' | 'Q' | 'R' | 'S' | 'T' | 'U' | 'V' | 'W' | 'X' | 'Y' | 'Z' ;
6960 portable_graph_gtr           = letter | digit | '!' | '"' | '#' | '$' | '%' | '&'
6961                               | "''" | '(' | ')' | '*' | '+' | '-' | '.' | '/' | ':' | ';' |
6962                               | '<' | '=' | '?' | '@' | '[' | ']' | '\`' | '^' | '_' | '_'
6963                               | '{' | '}' | ',' | '}' | '~~' ;
6964 portable_graph               = portable_graph_gtr | '>' ;
6965 portable_char                = portable_graph | ' ' | <NUL> | <ALERT>
6966                               | <BACKSPACE> | <TAB> | <CARRIAGE_RETURN>
6967                               | <NEWLINE> | <VERTICAL_TAB> | <FORM_FEED> ;

```

```
6968     octal_char          = escape_char      octdigit octdigit octdigit* ;
6969     hex_char            = escape_char 'x' hexdigit hexdigit hexdigit* ;
6970     decimal_char        = escape_char 'd' digit digit digit* ;
6971     number              = digit+ ;
6972     id_part              = letter | digit | '-' | '_' ;
6973     four_digit_hex_string = hex_upper hex_upper hex_upper hex_upper ;
6974     identifier           = letter id_part* ;
6975     simple_symbol         = space* '<' portable_graph_gtr+ '>' ;
6976     ucs_symbol            = space* '<U' four_digit_hex_string
6977                   [ four_digit_hex_string ] '>' ;
6978     quoted_string          = space* char_symbol* "''" ;
6979     quoted_nonempty_string = space* char_symbol+ "''" ;
6980     char_symbol            = char | charsymbol
6981                   | octal_char | hex_char | decimal_char ;
6982     elem_list              = elem+ ;
6983     elem                  = char_symbol | collsymbol | collelement ;
6984     symb_list              = collsymbol+ ;
6985     FDCC_set_name          = FDCC-name | "'' FDCC-name "'' ;
6986     copy_FDCC_set          = 'copy' FDCC_set_name EOL ;
6987     FDCC-name              = portable_graph+ ;
6988     semicolon              = space* ';' space* ;
6989     comma                 = space* ',' space* ;
6990     comment                = comment_char char* ;
6991
```

**Annex D**  
(informative)**Issues list**

This Technical Report presents a trial for defining a general mechanism to specify cultural conventions. Though its contents are developed in order to form a standard, it is decided to be a technical report in order to give information to public earlier.

The issues includes but are not limited to:

- 1) Whether the features which have their origin in ISO/IEC 9945-2 --POSIX Part 2 -- works well after its separation from ISO/IEC 9945-2 or not.
- 2) Whether it makes sense or not to have a default value, which may be considered as a recommendation, for each cultural convention item.
- 3) Whether each specification form fits for world-wide cultural variations or not.

The preparer of this report, ISO/IEC JTC1/SC22, expects the rapid progress of internationalization in the field of information technology will solve the above mentioned issues and this technical report will be used as a base for a new standard in near future.

**D.1 Comments from the Japanese member body**

Japan considered this document should not be published as an international standard for the following reasons:

- 1) It is not clear whether the features which have their origin in ISO/IEC 9945-2 -- POSIX Part 2 -- works well or not, after its separation from ISO/IEC 9945-2. Japan considers some mechanisms, e.g. "copy", will not work outside the POSIX environments.
- 2) It is not clear whether it makes sense or not to have a default value, which may be considered as a recommendation, for each cultural convention item. Japan is afraid that those default values are considered as Global Uniformity values -- see ISO/IEC TR 11017:1998 for details.
- 3) It is not clear whether each specification form fits for world-wide cultural variations or not.

**D.2 Comments from the U.S. member body**

The U.S. has repeatedly reviewed this document, and is firmly of the opinion that it should not be published as an international standard. The reasons for that assessment include, but are not limited to:

1. As an extension of the POSIX locale syntax (cf. ISO/IEC 9945-2), this document maintains the drawbacks of POSIX as a "specification method for cultural conventions", and in fact exacerbates the weaknesses of POSIX by conflating more, poorly justified LC\_XXX formal definitions into a monolithic FDCC-set construct. This was clearly done with a particular implementation model in mind, but does not follow, nor even seem to be

7044 particularly informed by best current practice in the internationalization of software.

7045  
7046 2. In an attempt to extend the POSIX LC\_CTYPE specification to cover the repertoire of  
7047 ISO/IEC 10646-1, this document blunders badly in asserting the cultural contextualization  
7048 of character properties for the UCS. The treatment of LC\_CTYPE as part of locales is an  
7049 artifact of POSIX architecture and results from the need to have a place to put locale  
7050 differences for case mapping. But by cloning other character properties having nothing to  
7051 do with case mapping into LC\_CTYPE, the net effect is to create a second source for  
7052 UCS character properties, with attendant dangers of divergence and errors, and with  
7053 inevitable difficulties of maintenance and versioning. The clear intent to force other ISO  
7054 standards to obtain their character property definitions from this document, instead of by  
7055 reference to the widely implemented UCS property tables published by the Unicode  
7056 Consortium, would lead to confusion and interoperability problems for character properties  
7057 -- just the opposite of what a standard should be doing.

7058  
7059 3. Each of the categories in the FDCC-set description has unaddressed problems and  
7060 limitations. Rather than being resolved during the development of this document, many of  
7061 these limitations were simply asserted to be "requirements" or necessary limitations. And  
7062 it appears to us that those are limitations of a particular envisioned implementation, rather  
7063 than following from the legitimate needs for specification of cultural conventions. Because  
7064 of this, implementers attempting to make use of the FDCC-set categories are immediately  
7065 faced with an unexplained host of problems and mismatches to the actual cultural  
7066 adaptability that they are trying to implement to meet customer needs for information  
7067 technology.

7068  
7069 4. In particular, the LC\_MONETARY extensions to deal with euro sign dual formatting  
7070 requirements seem to be an unnecessarily complicated scheme rolled out too late -- and do  
7071 not follow the approach already taken by production software to solve the problem.

7072  
7073 5. This document contains a completely unnecessary repertoire map definition intended to  
7074 promulgate a particularly bad collection of character mnemonic short strings. The U.S.  
7075 views these "mnemonics" as confusing and irrelevant. The need for short identifiers for  
7076 characters can be met by the standard short identifiers spelled out in ISO/IEC 10646,  
7077 which \*are\* in widespread use.

7078  
7079 6. There are numerous errors of detail in this document. While these could, in principle,  
7080 be addressed, many have not been. On that basis alone, it seems inadvisable to make the  
7081 document a standard.

7082  
7083 The U.S. does not share the optimistic assessment of the usefulness of this document as a  
7084 "trial" mechanism nor of the ease of addressing the issues mentioned here.

7086  
7087  
7088  
7089  
7090

	Annex E (informative)	Index	
7091	abbreviation	4.2	col_weight_max
7092	abday	4.7	collating-element
7093	abmon	4.7	collating statements
7094	absolute ellipses	4.3	collating-symbol
7095	address	4.2	collating element
7096	addresses	4.11	collating sequence
7097	addset	5.1	collating-element
7098	alpha	4.3.1	collating-symbol
7099	alt_digits	4.7	collation
7100	am_pm	4.7	combining
7101	application	4.2	combining_level3
7102	audience	4.2	comment_char
7103	blank	4.3.1	conformance
7104	block_separator	4.3.1	contact
7105	byte	3.1.1	continuation line
7106	cal_direction	4.7	control characters
7107	category	4.2	conversion_rate
7108	category names	4.1	copy
7109	category trailer	4.1	3.4.3, 4.2, 4.3.1, 4.4.2, 4.5, 4.6, 4.7, 4.8,
7110	category header	4.1	4.9,
7111	category body	4.1	4.10, 4.11, 4.12
7112	character	3.1.2	country_ab2
7113	character, graphic	4.3.1	country_ab3
7114	character, special	4.3.1	country_car
7115	character representation	4.1.1	country_isbn
7116	character, native digit	4.3.1	country_name
7117	character, hexadecimal digit	4.3.1	country_num
7118	character, multibyte	4.1.1	country_post
7119	character, decimal constant	4.1.1	cultural convention
7120	character, hexadecimal constant	4.1.1	currency_symbol
7121	character, space	4.3.1	d_fmt
7122	character, octal constant	4.1.1	d_t_fmt
7123	character, control	4.3.1	date field descriptors
7124	character, blank	4.3.1	date
7125	character, digit	4.3.1	day
7126	character, punctuation	4.3.1	decimal_point
7127	character, printable	3.1.10	default_missing
7128	character class	3.1.9	definitions
7129	character, coded	3.1.3	digit
7130	Character set rationale	B.2	ellipses
7131	charmap text	5.1	ellipses, absolute
7132	charmap	5, 4.1.4.4, 3.1.7	ellipses, symbolic
7133	charmap rationale	B.2	email
7134	class	4.3.1	equivalence class
7135	cntrl	4.3.1	era
7136	code_set_name	5.1	era_d_fmt
7137	coded character	3.1.3	era_year
			escape_char

7138	esqseq	5.1
7139	euro	B.1.3
7140	extended regular expression	4.8
7141	fax	4.2
7142	FDCC-set, definition	4.1
7143	FDCC-set	4f
7144	FDCC-set	3.1.6
7145	FDCC-set rationale	B.1
7146	first_weekday	4.7
7147	first_workday	4.7
7148	frac_digits	4.5
7149	graph	4.3.1
7150	graphic chracters	4.3.1
7151	grouping	4.6
7152	height	4.9
7153	include	4.3.2
7154	include	5.1
7155	include	4.3.2.2
7156	int_curr_symbol	4.5
7157	int_frac_digits	4.5
7158	int_n_cs_precedes	4.5
7159	int_n_sep_by_space	4.5
7160	int_n_sign_posn	4.5
7161	int_p_cs_precedes	4.5
7162	int_p_sep_by_space	4.5
7163	int_p_sign_posn	4.5
7164	int_prefix	4.12
7165	int_select	4.12
7166	keywords	4.1
7167	lang_ab	4.11
7168	lang_lib	4.11
7169	lang_name	4.11
7170	lang_term	4.11
7171	language	4.2
7172	LC_ADDRESS	4.11
7173	LC_ADDRESS rationale	B.1.10
7174	LC_COLLATE	4.4
7175	LC_COLLATE rationale	B.1.3
7176	LC_CTYPE	4.3
7177	LC_CTYPE rationale	B.1.2
7178	LC_IDENTIFICATION	4.2
7179	LC_IDENTIFICATION rationale	B.1.1
7180	LC_MESSAGES	4.8
7181	LC_MESSAGES rationale	B.1.7
7182	LC_MONETARY	4.5
7183	LC_MONETARY rationale	B.1.4
7184	LC_NAME	4.10
7185	LC_NAME rationale	B.1.9
7186	LC_NUMERIC	4.6
7187	LC_NUMERIC rationale	B.1.5
7188	LC_TELEPHONE	4.12
7189	LC_TELEPHONE rationale	B.1.11
	LC_TIME	4.7
	LC_TIME rationale	B.1.6
	LC_XLITERATE	4.9
	LC_XLITERATE rationale	B.1.8
	LC_X	4
	line continuation	4.1.4
	lower	4.3.1
	map	4.3.1
	mb_cur_max	5.1
	mb_cur_min	5.1
	messages	4.8
	modified date field descriptors	4.7.2

7190	mon	4.7	repertoire	6
7191	mon_decimal_point	4.5	repertoiremap	6, 3.1.8, 5.1, 4.1.4.3
7192	mon_grouping	4.5	revision	4.2
7193	mon_thousands_sep	4.5	scope	1
7194	monetary	4.5	section	4.4, 4.4.4
7195	multicharacter collating element	3.1.14	source	4.2
7196	n_cs_precedes	4.5	space	4.3.1
7197	n_sep_by_space	4.5	special characters	4.3.1
7198	n_sign_posn	4.5	symbol-equivalence	4.4, 4.4.7
7199	name formatting	4.10	symbolic ellipses	4.3, 5.1
7200	name_fmt	4.10	symbolic name	4.1.1
7201	name_gen	4.10	syntax format	3.2.1
7202	name_miss	4.10	t_fmt	4.7
7203	name_mr	4.10	t_fmt_ampm	4.7
7204	name_mrs	4.10	tel	4.2
7205	name_ms	4.10	tel_dom_fmt	4.12
7206	negative_sign	4.5	tel_int_fmt	4.12
7207	noexpr	4.8	telephone numbers	4.12
7208	notations	3.2	territory	4.2
7209	numeric	4.6	text file	3.1.4
7210	operands	4.1	thousands_sep	4.6
7211	order_end	4.4.9, 4.4	timezone	4.7
7212	order_start	4.4, 4.4.8	title	4.2
7213	outdigit	4.3.1	tolower	4.3.1
7214	p_cs_precedes	4.5	tosymmetric	4.3.1
7215	p_sep_by_space	4.5	toupper	4.3.1
7216	p_sign_posn	4.5	translit_ignore	4.9
7217	paper format	4.9	transliteration	4.9
7218	portable character set	3.2.4	transliteration statements	4.9.1
7219	positive_sign	4.5	upper	4.3.1
7220	POSIX	1	valid_from	4.5
7221	POSIX differences	A	valid_to	4.5
7222	POSIX conformance	4.2	visible glyph portable characters	3.2.4
7223	postal addresses	4.11	week	4.7
7224	postal_fmt	4.11	white space	3.1.11
7225	pre-category statements	4.1.4	width	4.9
7226	print	4.3.1	xdigit	4.3.1
7227	printable character	3.1.10	yesexpr	4.8
7228	punct	4.3.1		
7229	punctuation characters	4.3.1		
7230	redefine	4.3.2		
7231	references	2		
7232	reorder-section-end	4.4.13		
7233	reorder-section-after	4.4.12		
7234	reorder-section-after	4.4		
7235	reorder-after	4.4		
7236	reorder-end	4.4		
7237	reorder-section-end	4.4		
7238	reorder-after	4.4.10		
7239	reorder-end	4.4.11		
7240	reorder-after rationale	B.1.2.1		
7241	repertoire rationale	B.3		

**BIBLIOGRAPHY**

- 7242  
7243  
7244 The following specifications are considered relevant to this Technical Report, in addition  
7245 to the normative references.  
7246  
7247 CEPT, CEPT-MAILCODE, *Country code for mail*.  
7248  
7249 ISO 646, *Information technology - ISO 7-bit coded character set for information inter-*  
7250 *change*.  
7251  
7252 ISO/IEC 9899, *Information technology - Programming language C*.  
7253  
7254 ISO/IEC 14977, *Information technology - Syntactic metalanguage - Extended BNF*.  
7255  
7256 The Unicode Consortium: *The Unicode Standard, Version 2.0*, Addison Wesley  
7257 Developers Press, July 1996. ISBN 0-201-48345-9.  
7258  
7259 IBM: *National Language Design Guide Volume 2 - National Language Support Reference*  
7260 *Manual*, IBM SE09-8002-03, August 1994.  
7261  
7262 STRÍ: *Nordic Cultural Requirements on Information Technology (Summary report)*, STRÍ  
7263 TS3, Libris, Reykjavík, Iceland 1992. ISBN 9979-9004-3-1.  
7264  
7265