

ANSI®
Z39.47-1985
ISSN: 8756-0860

American National Standard
for Information Sciences –
**Extended Latin Alphabet Coded
Character Set for Bibliographic Use**

Secretariat
Council of National Library and Information Associations

Approved September 4, 1984

American National Standards Institute, Inc

Abstract

This standard establishes computer codes for an extended Latin alphabet character set to be used in bibliographic work when handling non-English items. This standard establishes the 7-bit and 8-bit code values.

Foreword (This Foreword is not part of American National Standard Z39.47-1985.)

This standard for the extended Latin alphabet coded character set for bibliographic use was developed by Subcommittee N on Coded Character Sets for Bibliographic Information Interchange of the American National Standards Committee on Library and Information Sciences and Related Publishing Practices, Z39, now the National Information Standards Organization (Z39).

The codes given in this standard may be identified by the use of the notation ANSEL. The notation ANSEL should be taken to mean the codes prescribed by the latest edition of this standard. To explicitly designate a particular edition, the last two digits of the year of the issue may be appended, as in ANSEL 85.

The standard establishes both the 7-bit and the 8-bit code values for the computer codes for characters used in bibliographic work when handling non-English items. The characters included in the codes have been selected because they are the ones needed to fully record bibliographic citations in many Latin alphabet languages and non-Latin languages transliterated into Latin alphabet characters.

The greater part of the extended Latin alphabet coded character set in this standard has been used for bibliographic work in the library community for over 15 years. Thus, it predates extended character set standardization that has recently been undertaken by the American National Standards Committee on Information Processing Systems, X3, and by the International Organization for Standardization Technical Committees TC 46 on Documentation and TC 97 on Information Processing Systems. The characters in column 4 (7-bit) or 12 (8-bit) are additions to this library set that have been identified by Subcommittee N as valuable.

Subcommittee N is continuing to work to develop the standards that will make it possible to access the traditional library character sets in accordance with the applicable code extension standards. It is anticipated that a standard for a Greek alphabet coded character set will follow shortly.

Suggestions for improvement of this standard will be welcome. They should be sent to the Council of National Library and Information Associations, 461 West Lancaster Avenue, Haverford, PA 19041.

This standard was processed and approved for submittal to ANSI by Committee Z39. Committee approval of the standard does not necessarily imply that all committee members voted for its approval. At the time it approved this standard, the Z39 Committee had the following members:

EXECUTIVE COUNCIL

Patricia W. Berger, Chair

(Chief, Information Resources and Services Division, National Bureau of Standards)

Sandra K. Paul, Vice-Chair

(SKP Associates, Inc)

Councilors: Libraries

Linda K. Bartley

(Library of Congress)

Sally H. McCallum

(Library of Congress)

Larry X. Besant

(Linda Hall Library)

Councilors: Information Services

James E. Rush

(James E. Rush Associates, Inc)

W. Theodore Brandhorst

(ERIC Processing and Reference Facility)

M. Lynne Neufeld

(National Federation of Abstracting and Information Services)

Councilors: Publishing

Karl F. Heumann

(Federation of American Societies for Experimental Biology)

Seldon W. Terrant

(American Chemical Society)

Executive Director

Patricia R. Harris

Secretariat Representative

John T. Corrigan

(Council of National Library and Information Associations)

<i>Organization Represented</i>	<i>Name of Representative</i>
American Association of Law Libraries	Robert Oakley
American Chemical Society	James L. Wood
	Robert S. Tannehill, Jr (Alt)
American Library Association.	Susan H. Vita
	Janice E. Anderson (Alt)
American Psychological Society.	Ann I. Mahoney
American Society for Information Science	David Liston
American Society of Indexers.	Hans H. Wellisch
American Theological Library Association.	Warren Kissinger
Association of American Publishers.	Sandra K. Paul
Association of Information and Dissemination Centers (ASIDIC)	Maureen Kelly
Association of Jewish Libraries	Herbert C. Zafren
Association of Research Libraries.	Shirley Echelman
Book Manufacturers' Institute.	Douglas Horner
R. R. Bowker Company, Inc.	Emery Koltay
Colorado Alliance of Research Libraries (CARL)	Ward Shaw
Cooperative College Library Center (CCLC).	Hillis D. Davis
Council of Biology Editors.	Grace Gurtowski
Council of National Library and Information Associations (CNLIA).	John T. Corrigan
Engineering Societies Library	S. K. Cabeen
F. W. Faxon Company, Inc	Dan Tonkery
Indiana Cooperative Library Services Authority (INCOLSA)	Barbara Markuson
	Janice Alexander (Alt)
Information Industry Association.	Albert L. Batik
Library Binding Institute.	Albert Leitschuh
Library of Congress.	Joseph H. Howard
Medical Library Association.	Nancy N. Woelfl
Music Library Association	Kathleen Moretto
National Agricultural Library	Joseph H. Howard
	Samuel T. Waters (Alt)
National Archives and Records Service.	Alan Calmes
National Bureau of Standards, Information Resources and Services Division	Claire Tozier
	Diane Cunningham (Alt)
National Commission on Libraries and Information Science (NCLIS)	Toni C. Bearman
	Dorothy P. Gray (Alt)
National Federation of Abstracting and Information Services (NFAIS)	M. Lynne Neufeld
National Library of Medicine	Lois A. Colaianni
	Joseph Gantner (Alt)
OCLC, Inc	Randolph Call
	Mary Ellen Jacob (Alt)
OHIONET	Ronald E. Diener

<i>Organization Represented</i>	<i>Name of Representative</i>
PALINET	James Schoenung
Pittsburgh Regional Library Center (PRLC)	Scott Bruntjen
Research Libraries Group, Inc.	Tina Kass
	Wayne Davison (Alt)
Society for Scholarly Publishing	Seldon W. Terrant
	Robert W. Frase (Alt)
Society for Technical Communication	Ronald M. Field
Special Libraries Association	R. L. Funkhouser
SUNY/OCLC Network	Glyn T. Evans
	Teresa Strozik (Alt)
U.S. Department of Commerce, National Technical Information Service (NTIS)	Kenyon Rosenberg
U.S. Department of Commerce, Printing and Publishing Division	William Lofquist
U.S. Department of Defense	Dorothy A. Fisk
U.S. Department of Education, Office of Libraries and Learning Resources	Robert Klassen
U.S. Department of Energy, Technical Information Center	James D. Cape
	Lee Thompson (Alt)

Subcommittee N, which developed this standard, had the following members:

Sally H. McCallum, Co-Chair (Library of Congress)	James E. Agenbroad (Library of Congress)
Charles Payne, Co-Chair (University of Chicago Library)	Nelson Chou (Rutgers University Library)
	Wayne Davison (RLG, Inc)
	Susan Olson (OCLC, Inc)
	Gerald Reisz (Chemical Abstracts Service, Research and Development)
	David L. Weisbrod (Yale University)

Contents	SECTION	PAGE
	1. Scope, Purpose, and Application	6
	1.1 Scope	6
	1.2 Purpose	6
	1.3 Application	6
	2. Referenced Standards.	7
	2.1 American National Standards.	7
	2.2 ISO Standard	7
	3. Definitions.	7
	4. Implementation	7
	5. Code Tables for the Extended Latin Alphabet Coded Character Set	8
	5.1 7-Bit Code Table	8
	5.2 8-Bit Code Table	8
	6. Legend	11
	6.1 ANSEL Spacing Graphic Characters	11
	6.2 ANSEL Nonspacing Graphic Characters.	12
	6.3 ASCII Control Characters	13
	6.4 ASCII Graphic Characters	14
	Figures	
	Figure 1 7-Bit Code Table	9
	Figure 2 8-Bit Code Table	10
	Appendixes	
	Appendix A Languages Using ANSEL Character Modifiers or Special Characters	15
	Tables	
	Table A1 Latin Alphabet Languages.	15
	Table A2 Transliterated Non-Latin Alphabet Languages.	17
	Appendix B ANSEL Character Modifiers and Special Characters with Languages of Occurrence	20
	Tables	
	Table B1 Character Modifiers	20
	Table B2 Special Characters	23

American National Standard for Information Sciences –

Extended Latin Alphabet Coded Character Set for Bibliographic Use

1. Scope, Purpose, and Application

1.1 Scope. This standard specifies 63 graphic characters contained in a 94-byte set that can be invoked in a 7-bit or 8-bit environment. They are intended for use with the 128 graphic and control characters of ASCII, the American National Standard Code for Information Interchange, ANSI X3.4-1977, and are therefore fully compatible with the 7-bit coded character set defined in that standard. (The relationship and use of multiple character sets is described in the American National Standard Code Extension Techniques for Use with the 7-Bit Coded Character Set of American National Standard Code for Information Interchange, ANSI X3.41-1974.)

This standard consists of code tables and a legend giving the name and an example for each of the extended Latin graphic characters.

1.2 Purpose. The character set in this standard is intended for the interchange of bibliographic information among data processing systems and within message transmission systems. It is suitable for bibliographic citations, including their annotations, in the Latin alphabet.

1.3 Application. The character set in this standard is intended to handle information in the following languages, among others:

Afrikaans	French
Albanian	German
Anglo-Saxon	Hawaiian
Catalan	Hungarian
Czech	Icelandic (Modern)
Danish	Indonesian
Dutch	Italian
English	Latvian
Esperanto	Lithuanian
Estonian	Navaho
Faroese	Norwegian
Finnish	Polish

Portuguese	Swedish
Romanian	Tagalog
Serbo-Croatian (Latin)	Turkish (Modern)
Slovak	Vietnamese
Slovene	Wendic
Spanish	

It is also intended to handle romanized forms of the following languages, among others:

Amharic	Malayalam
Arabic	Marathi
Armenian	Manipuri
Assamese	Mewari
Belorussian	Nepali
Bengali	Oriya
Braj	Pahari
Bulgarian	Pali
Burmese	Panjabi
Chinese	Persian
Church Slavic	Prakrit
Dogri	Pushto
Georgian	Rajasthani
Greek	Russian
Gujarati	Sanskrit
Hebrew	Serbo-Croatian (Cyrillic)
Hindi	Sindhi
Japanese	Sinhalese
Kannada	Tamil
Khmer	Telugu
Konkani	Thai
Korean	Tibetan
Lahnda	Ukrainian
Lao	Urdu
Macedonian	Yiddish
Maithili	

Appendix A contains two tables showing the character modifiers and special characters coded in this standard as they are used for each language. Appendix B contains two tables showing the languages that use each character modifier or special character.

2. Referenced Standards

2.1 American National Standards. This standard is intended for use in conjunction with the following American National Standards. When these standards are superseded by a revision approved by the American National Standards Institute, Inc, the revision shall apply:

ANSI X3.4-1977, American National Standard Code for Information Interchange

ANSI X3.41-1974, American National Standard Code Extension Techniques for Use with the 7-Bit Coded Character Set of American National Standard Code for Information Interchange

ANSI Z39.11-1972, American National Standard System for the Romanization of Japanese

ANSI Z39.12-1972, American National Standard System for the Romanization of Arabic

ANSI Z39.24-1976, American National Standard System for the Romanization of Slavic Cyrillic Characters

ANSI Z39.25-1975, American National Standard Romanization of Hebrew

ANSI Z39.35-1979, American National Standard System for the Romanization of Lao, Khmer, and Pali

ANSI Z39.37-1979, American National Standard System for the Romanization of Armenian

2.2 ISO Standard. This new standard is intended for use in conjunction with Data Processing – Procedure for Registration of Escape Sequences, ISO 2375-1980.¹

3. Definitions

character modifiers (diacritics). A mark, point, or sign used with alphabetic graphic characters to distinguish them in form or sound.

control character. A character whose occurrence in a particular context initiates, modifies, or stops an action that affects the recording, processing, transmission, or interpretation of data.

graphic character. A character, other than a control character, that has a visual representation normally handwritten, printed, or displayed.

nonspacing graphic character. A graphic character whose use is not followed by the forward movement of the output device. For the purposes of this standard, the term includes character modifiers.

spacing graphic character. A graphic character whose use is followed by the forward movement of the output device to the next character position. For the purposes of this standard, the term includes special characters, special symbols, and punctuation marks.

punctuation marks. A mark that indicates the structure of sentence parts for clearness; for example, . : ; .

special character. An alphabetic character or other spacing graphic character; for example, æ ' ð .

special symbols. A conventional sign used in place of words or word groups; for example, © % ® .

4. Implementation

4.1 The implementation of this American National Standard is in accordance with the provisions of ANSI X3.41-1974. The 7-bit supplementary set is identified by escape sequences assigned by the ISO Registration Authority in accordance with procedures given in ISO 2375-1980.

4.2 The unassigned positions in the code tables shall not be used.

4.3 Character modifiers, which are always used in conjunction with other characters, appear in columns 6 and 7 of the 7-bit extended Latin alphabet coded character set (columns 14 and 15 in an 8-bit environment). All characters in these two columns are designated nonspacing graphic characters. Thus, the Backspace character, column 0/row 8 (col/row, 0/8), is not used with these character modifiers. In a character string, these nonspacing characters shall precede the character that they modify. When a character requires multiple character modifiers, they are to be entered in the order in which they appear, reading left to right or top to bottom.

4.4 The following character modifiers appear in the ASCII set as spacing characters and are defined as nonspacing characters in the extended Latin alphabet coded character set. The nonspacing form shall be used.

Name	ASCII Col/Row	ANSEL	
		7-Bit Col/Row	8-Bit Col/Row
circumflex	5/14	6/3	14/3
underline	5/15	7/6	15/6
tilde	7/14	6/4	14/4

¹ Published by the International Organization for Standardization (ISO) and available from the American National Standards Institute, 1430 Broadway, New York, NY 10018.

4.5 The following ASCII characters are indicated as having alternative use as character modifiers in ANSI X3.4-1977. These ASCII characters shall not be used as character modifiers (requiring the Backspace) but the corresponding nonspace characters in the extended Latin alphabet coded character set shall be used.

Name (Alternative Name)	ASCII Col/Row	ANSEL	
		7-Bit Col/Row	8-Bit Col/Row
quotation mark (diaeresis)	2/2	6/8	14/8
apostrophe (acute accent)	2/7	6/2	14/2
comma (cedilla)	2/12	7/0	15/0
opening single quotation mark (grave accent)	6/0	6/1	14/1

5. Code Tables for the Extended Latin Alphabet Coded Character Set

5.1 7-Bit Code Table. The extended Latin alphabet coded character set in a 7-bit environment shall be as shown in Figure 1, 7-Bit Code Table. The legend is in Section 6.

5.2 8-Bit Code Table. The extended Latin alphabet coded character set in an 8-bit environment shall be as shown in Figure 2, 8-Bit Code Table.

In columns 0 through 7, the table shows the characters of the 7-bit coded character set ASCII adapted for 8 bits. The extended Latin alphabet coded character set is shown in columns 10 through 15. The legend is in Section 6.

		B I T S							
		0	1	2	3	4	5	6	7
0000	0				'	°		?	„
0001	1			ℓ	ℓ	/		`	˘
0010	2			ø	ø	®		'	•
0011	3			Ɔ	ǎ	©		^	..
0100	4			Ɔ	Ɔ	£		~	°
0101	5			Æ	æ	¿		-	=
0110	6			œ	œ	ı		ı	-
0111	7			'	"			•	˘
1000	8			•	ı			"	˘
1001	9			b	£			˘	˘
1010	10			®	ǎ			°	˘
1011	11			±				˘	˘
1100	12			σ	σ			˘	
1101	13			ı	ı			'	
1110	14			'				"	'
1111	15							ı	





- 4321
B I T S
-  Reserved for control characters.
-  Reserved for future standardization.
-  Corners (Reserved).


Figure 1
7-Bit Code Table

4321

BITS

***Redefined in the extended Latin alphabet coded character set.**

 Reserved for control characters.

 Reserved for future standardization.

Corners (Reserved).

Figure 2 8-Bit Code Table

6. Legend

NOTE: The legend for ASCII characters in 6.3 and 6.4 is adapted from ANSI X3.4-1977. For complete information on the use of ASCII refer to the latest edition of that standard.

6.1 ANSEL Spacing Graphic Characters

7-Bit Col/Row	8-Bit Col/Row	Graphic	Name	Example of Use
2/1	10/1	ℒ	slash L — uppercase	ℒódz
2/2	10/2	Ø	slash O — uppercase	Øst
2/3	10/3	Ð	slash D — uppercase	Ðuro
2/4	10/4	Þ	thorn — uppercase	Þann
2/5	10/5	Æ	ligature AE — uppercase	Ægir
2/6	10/6	Œ	ligature OE — uppercase	Œuvre
2/7	10/7	ˆ	mîâgkiř znak	Fakul'tet
2/8	10/8	•	middle dot	novel•la
2/9	10/9	♭	musical flat	B♭
2/10	10/10	®	patent mark	ABC®
2/11	10/11	±	plus or minus	A ± B
2/12	10/12	Œ	hook O — uppercase	BŒ
2/13	10/13	Ů	hook U — uppercase	XUŮ
2/14	10/14	ʾ	alif	Un'yusho
3/0	11/0	ˆ	ˆayn	fa'ıl
3/1	11/1	ℓ	slash l — lowercase	rozbiℓ
3/2	11/2	ø	slash o — lowercase	høj
3/3	11/3	đ	slash d — lowercase	đavola
3/4	11/4	þ	thorn — lowercase	þann
3/5	11/5	æ	ligature æ — lowercase	skæg
3/6	11/6	œ	ligature œ — lowercase	œuvre
3/7	11/7	ˆ	tvěrdyř znak	ob'îâvlenie
3/8	11/8	ı	dotless i — lowercase	masalı
3/9	11/9	£	British pound	£5.00
3/10	11/10	ð	eth	verður
3/12	11/12	σ	hook o — lowercase	Sσ
3/13	11/13	υ	hook u — lowercase	Tυ Duc
4/0	12/0	°	degree sign	10°C
4/1	12/1	ℓ	script l ²	25 ℓ.
4/2	12/2	®	phonograph record copyright mark	Decca ®
4/3	12/3	©	copyright mark	© 1974
4/4	12/4	♯	musical sharp	D♯
4/5	12/5	¿	inverted question mark	¿Qué?
4/6	12/6	¡	inverted exclamation mark	¡Esta!

²In bibliographic work, the script l, ℓ, is commonly used as an abbreviation for the term "leaves." It shall not be used as a symbol for the unit of measure "liter."

6.2 ANSEL Nonspacing Graphic Characters

7-Bit Col/Row	8-Bit Col/Row	Graphic	Name	Example of Use
6/0	14/0	˘	low rising tone mark	cùl
6/1	14/1	`	grave accent	règle
6/2	14/2	´	acute accent	está
6/3	14/3	^	circumflex accent	même
6/4	14/4	~	tilde	niño
6/5	14/5	—	macron	gājējs
6/6	14/6	˘	breve	altă
6/7	14/7	˙	dot above	žaba
6/8	14/8	¨	umlaut (diaeresis)	öppna
6/9	14/9	ˇ	háček	vždy
6/10	14/10	°	circle above (angstrom)	hår
6/11	14/11	ˆ	ligature, left half	akademiiā
6/12	14/12	˜	ligature, right half	akademiiā
6/13	14/13	¸	high comma, off center	rozdeľovac
6/14	14/14	˝	double acute accent	idoszaki
6/15	14/15	◌̣	candrabindu	Aliiëv
7/0	15/0	¸	cedilla	ça
7/1	15/1	ˆ	right hook	vieta
7/2	15/2	˙	dot below	teḍa
7/3	15/3	¨	double dot below	khutbah
7/4	15/4	°	circle below	Maharṣicaritāmṛtam
7/5	15/5	≡	double underscore	Ghulam
7/6	15/6	—	underscore	samar
7/7	15/7	ˆ	left hook	darziṇa
7/8	15/8	˘	right cedilla	khọng
7/9	15/9	˘	half circle below (upadhmaniya)	ḥumantuṣ
7/10	15/10	ˆ	double tilde, left half	ñgalan
7/11	15/11	˜	double tilde, right half	ñgalan
7/14	15/14	¸	high comma, centered	geotermika

6.3 ASCII Control Characters

Col/ Row	Mnemonic and Meaning	Col/ Row	Mnemonic and Meaning
0/0	NUL Null	1/0	DLE Data Link Escape
0/1	SOH Start of Heading	1/1	DC1 Device Control 1
0/2	STX Start of Text	1/2	DC2 Device Control 2
0/3	ETX End of Text	1/3	DC3 Device Control 3
0/4	EOT End of Transmission	1/4	DC4 Device Control 4
0/5	ENQ Enquiry	1/5	NAK Negative Acknowledge
0/6	ACK Acknowledge	1/6	SYN Synchronous Idle
0/7	BEL Bell	1/7	ETB End of Transmission Block
0/8	BS Backspace	1/8	CAN Cancel
0/9	HT Horizontal Tabulation	1/9	EM End of Medium
0/10	LF Line Feed	1/10	SUB Substitute
0/11	VT Vertical Tabulation	1/11	ESC Escape
0/12	FF Form Feed	1/12	FS File Separator
0/13	CR Carriage Return	1/13	GS Group Separator
0/14	SO Shift Out	1/14	RS Record Separator
0/15	SI Shift In	1/15	US Unit Separator
		7/15	DEL Delete

6.4 ASCII Graphic Characters

Col/Row	Graphic	Name
2/0	SP	Space (Normally Nonprinting)
2/1	!	Exclamation Point
2/2	"	Quotation Marks (Diaeresis)
2/3	#	Number Sign
2/4	\$	Dollar Sign
2/5	%	Percent Sign
2/6	&	Ampersand
2/7	'	Apostrophe (Closing Single Quotation Mark; Acute Accent)
2/8	(Opening Parenthesis
2/9)	Closing Parenthesis
2/10	*	Asterisk
2/11	+	Plus
2/12	,	Comma (Cedilla)
2/13	-	Hyphen (Minus)
2/14	.	Period (Decimal Point)
2/15	/	Slant
3/0 to 3/9	0 . . . 9	Digits 0 through 9
3/10	:	Colon
3/11	;	Semicolon
3/12	<	Less Than
3/13	=	Equals
3/14	>	Greater Than
3/15	?	Question Mark
4/0	@	Commercial At
4/1 to 5/10	A . . . Z	Uppercase Latin Letters A through Z
5/11	[Opening Bracket
5/12	\	Reverse Slant
5/13]	Closing Bracket
5/14	^	Circumflex
5/15	_	Underline
6/0	`	Opening Single Quotation Mark (Grave Accent)
6/1 to 7/10	a . . . z	Lowercase Latin letters a through z
7/11	{	Opening Brace
7/12		Vertical Line
7/13	}	Closing Brace
7/14	~	Tilde

Appendixes (These Appendixes are not part of American National Standard Z39.47-1985, but are included for information only.)

Appendix A

Languages Using ANSEL Character Modifiers or Special Characters

In the following tables, the Latin alphabet languages and transliterated non-Latin alphabet languages are listed with ANSEL character modifiers and special characters found in each language. Only lowercase letters are shown unless the uppercase form would not be obvious, in which case, it is shown in parentheses. In cases in which character modifiers are used as pronunciation signs (as in Afrikaans or Vietnamese), only a few of the possible character modifiers with alphabetic character combinations are shown.

The transliteration characters are those required for the ANSI Z39 romanization tables (listed in Section 2 of this standard) and the Library of Congress romanization tables.³ For the romanization of Hebrew, the Library of Congress table and Tables 1 and 2 of ANSI Z39.25-1975 were used.

³The romanization tables adopted by the American Library Association and the Library of Congress. For further information contact the Office for Descriptive Cataloging Policy, Processing Services, Library of Congress, Washington, DC 20540.

Table A1
Latin Alphabet Languages

Language	Character Modifiers	Special Characters
Afrikaans	ê î ô û ë ï è ò (etc)	
Albanian	á é ó ú í â ê ë ç	
Anglo-Saxon	ā āc ē ī ō ū ę ȝ	æ œ þ ð
Catalan	à è ò é í ó ú ï ü ç	•
Czech	á é í ó ú ý č ď ě ě ř š ř ž d' (Ď) t' (Ť) ů ý ŷ	
Danish	å ģ	æ ø
Dutch	â ê ë ï á é	
Esperanto	ĉ ĝ ĥ ĵ ŝ ŭ	
Estonian	ä ö ü š ž õ	
Faroese	á í ó ú ý	ø æ ǿ
Finnish	ä å	
French	à è ù â ê î ô û é ç ë ü ï	œ
German	ä ö ü	
Hawaiian	á é í ó ú ā ē ī ō ū	‘
Hungarian	á é í ó ú ő ű ö ü á	
Icelandic (Modern)	á é í ó ú ý ö	æ ð þ

(Continued on next page)

APPENDIX

Table A1 (Continued)

Language	Character Modifiers	Special Characters
Indonesian		² ²
Italian	à è ì ò ù	
Latvian	ā ģ ī ū ķ ļ ņ š ž ģ (Ģ)	
Lithuanian	ą ę į ū č š ž ė ū	
Navaho	ê ñ ħ ą į á ê í ġ ā	ɿ æ ʻ
Norwegian	å ä ö	æ ø
Polish	ą ę í ñ ó ś ź ę ż y	ɿ
Portuguese	ā ē ī ō ū á à ç ô ü	
Romanian	â î ș ç ț ă	
Serbo-Croatian (Latin)	č š ž ć í ʹ	đ
Slovak	á é í ó ú ý í ě č ě ň ř š ž ď ě ď (Ď) ľ (Ľ) ť (Ť) ň ä ô	
Slovene	č ě š ž á ô vʹ	
Spanish	á é í ó ú ñ ü	
Swedish	ä ö ʀ å	
Tagalog	á à â ñg	
Turkish (Modern)	ö ü ç ş â û ğ ĩ (İ) î (Î)	ı (I) ʻ
Vietnamese	â ê ô ă à ò á ă ă ă ư } (etc) à ă ă ă ơ ơ ơ ơ ă	đ ơ ư
Wendic	ć ń ó ń ó ń ó ń ó ń ó ń ó č ě š ž ľ (Ľ) ò	ɿ

NOTE: The superscript two, ², is shown here for reference purposes only. It does not appear in the extended Latin alphabet coded character set as described in this standard.

Table A2

(Continued on next page)

Table A2 (Continued)

Language	Character Modifiers	Special Characters
Lahnda	ā ī ū ṭ ḍ ṇ ṛ ṁ ḱ ḥ ḡ ṇ ṇ ṁ	
Lao	ā ī ū ē ā̄ ō ā̄ ṽ ṽ	æ œ ' ‘
Macedonian	š č ž k g c	đ
Maithili	ā ī ū ṛ ṛ ṛ ṛ ṛ ṛ ṛ ṛ ṛ ṛ ṭ ṣ ḥ ḛ ḅ ḅ ḅ ḅ ḅ ḅ ḅ ḅ ś ḱ ḥ ḡ ḡ ḥ	,
Malayalam	ā ī ū ē ō ṛ ṛ ṛ ṛ ṛ ṛ ṛ ṛ ḍ ṇ ṣ ṁ ḥ ḁ ṇ ṇ ś ṛ	,
Manipuri	ā ī ū ṛ ṛ ṛ ṛ ṛ ṛ ṛ ṛ ṛ ṛ ṽ ṇ ś ṇ ṁ	,
Marathi	ā ī ū ṛ ṛ ṛ ṛ ṛ ṛ ṛ ṛ ṛ ṛ ō ṇ ṇ ś ā ī ū ē ō ā ī ū ē ō	,
Mewari	ā ī ū ṛ ṛ ṛ ṛ ṛ ṛ ṛ ṛ ṛ ṛ ṭ ṣ ḥ ḛ ḅ ḅ ḅ ḅ ḅ ḅ ḅ ḅ ś ḱ ḥ ḡ ḡ ḥ	,
Nepali	ā ī ū ṛ ṛ ṛ ṛ ṛ ṛ ṛ ṛ ṛ ṛ ṭ ṣ ḥ ḛ ḅ ḅ ḅ ḅ ḅ ḅ ḅ ḅ ś ḱ ḥ ḡ ḡ ḥ	,
Oriya	ā ī ū ṛ ṛ ṛ ṛ ṛ ṛ ṛ ṛ ṛ ṛ ḥ ṇ ṽ ṇ ṁ ṇ ś	,
Pahari	ā ī ū ṛ ṛ ṛ ṛ ṛ ṛ ṛ ṛ ṛ ṛ ṭ ṣ ḥ ḛ ḅ ḅ ḅ ḅ ḅ ḅ ḅ ḅ ś ḱ ḥ ḡ ḡ ḥ	,
Pali	ā ī ū ṇ ṭ ḍ ṇ ṛ ṁ ḥ ṇ ṇ ī	
Panjabi	ā ī ū ṭ ḍ ṇ ṛ ṁ ṇ ṇ ṁ ḱ ḥ ḡ	
Persian	ṣ ṣ ḥ ṣ ṣ ṣ ṣ ṣ ṣ ṣ ṣ ā ī ū á	, ‘ ,
Prakrit	ā ī ū ṛ ṛ ṛ ṛ ṛ ṛ ṛ ṛ ṛ ṛ ṇ ṇ ṇ ḥ ś ṁ ḥ	,
Pushto	ā ū ī ṭ ḥ ḍ ṛ ṣ ṣ ṇ ḁ ṣ ṭ ṣ ṣ ṣ ṣ ṣ ṣ	, ‘ ,
Rajasthani	ā ī ū ṛ ṛ ṛ ṛ ṛ ṛ ṛ ṛ ṛ ṛ ṭ ṣ ḥ ḛ ḅ ḅ ḅ ḅ ḅ ḅ ḅ ḅ ś ḱ ḥ ḡ ḡ ḥ	,

(Continued on next page)

Appendix B

ANSEL Character Modifiers and Special Characters with Languages of Occurrence

In the following tables, ANSEL character modifiers and special characters are listed with the languages in which each occurs. Double character modifiers that are sometimes used in Vietnamese and other languages are not given a separate entry.

The transliterations take into consideration the tables that are indicated in Appendix A.

Table B1
Character Modifiers

Name	Examples	Languages
acute accent	á é í ó ú í ñ ó p f s ú w v z	Albanian, Catalan, Czech, Dutch, Faroese, French, Hawaiian, Hungarian, Icelandic (Modern), Navaho, Polish, Portuguese, Serbo-Croatian (Latin), Slovak, Slovene, Spanish, Tagalog, Vietnamese, Wendic Transliterated: Church Slavic, Gujarati, Hindi, Kannada, Khmer, Konkani, Maithili, Malayalam, Manipuri, Marathi, Mewari, Nepali, Oriya, Pahari, Persian, Prakrit, Pushto, Rajasthani, Sanskrit, Serbo-Croatian (Cyrillic), Sindhi, Sinhalese, Tamil, Telugu, Tibetan, Urdu, Yiddish
breve	ă ě ĝ ħ ĩ Ĺ ŏ ů	Esperanto, Navaho, Romanian, Turkish (Modern), Vietnamese Transliterated: Belorussian, Braj, Bulgarian, Chinese, Church Slavic, Dogri, Georgian, Hindi, Khmer, Korean, Lahnda, Maithili, Mewari, Nepali, Pahari, Panjabi, Rajasthani, Russian, Sindhi, Ukrainian
candrabindu	ṁ ṅ ṛ	Transliterated: Assamese, Bengali, Braj, Bulgarian, Hindi, Maithili, Manipuri, Mewari, Nepali, Oriya, Pahari, Prakrit, Rajasthani, Sanskrit, Sindhi, Telugu, Tibetan
cedilla	ç š	Albanian, Catalan, French, Portuguese, Turkish (Modern)
circle above (angstrom)	å Ů	Czech, Danish, Norwegian, Slovak, Swedish
circle below	ł ɾ	Transliterated: Assamese, Bengali, Braj, Gujarati, Hindi, Kannada, Khmer, Konkani, Maithili, Malayalam, Manipuri, Marathi, Mewari, Nepali, Oriya, Pahari, Prakrit, Rajasthani, Sanskrit, Sinhalese, Telugu, Tibetan

(Continued on next page)

Table B1 (Continued)

Name	Examples	Languages
circumflex accent	â ê ĝ ĥ ĵ ô ŝ û ŵ ŷ	Afrikaans, Albanian, Dutch, Esperanto, French, Navaho, Portuguese, Romanian, Slovak, Slovene, Tagalog, Turkish (Modern), Vietnamese Transliterated: Amharic, Braj, Chinese, Gujarati, Hindi, Khmer, Maithili, Marathi, Mewari, Nepali, Pahari, Rajasthani, Sindhi, Sinhalese, Telugu
dot above	č ċ ě ě ĝ ħ ĩ ĵ ṁ ṅ ṣ ṭ ṭ ṭ ṭ ŵ ž	Navaho, Turkish Transliterated: Amharic, Assamese, Belorussian, Bengali, Braj, Church Slavic, Dogri, Georgian, Gujarati, Hebrew, Hindi, Kannada, Khmer, Konkani, Lahnda, Lithuanian, Maithili, Malayalam, Manipuri, Marathi, Mewari, Nepali, Oriya, Pahari, Pali, Panjabi, Prakrit, Pushto, Rajasthani, Russian, Sanskrit, Sindhi, Sinhalese, Tamil, Telugu, Tibetan
dot below	ḁ ḃ ḅ ḇ ḥ ḫ ṗ ṑ ṓ ṕ ṓ ṓ ṕ ṑ ṓ ṕ ṓ ṓ ṕ ṑ ṓ ṕ ṓ ṓ	Vietnamese Transliterated: Amharic, Arabic, Assamese, Bengali, Braj, Dogri, Georgian, Gujarati, Hebrew, Hindi, Kannada, Khmer, Konkani, Maithili, Malayalam, Manipuri, Marathi, Mewari, Nepali, Oriya, Pahari, Panjabi, Persian, Prakrit, Pushto, Rajasthani, Sanskrit, Sindhi, Sinhalese, Tamil, Telugu, Tibetan, Urdu, Yiddish
double acute accent	ű ő ű	Hungarian
double dot below	Ḃ Ḅ Ḇ Ḉ Ḭ Ḱ Ḳ	Transliterated: Braj, Hindi, Kannada, Konkani, Maithili, Mewari, Nepali, Pahari, Persian, Pushto, Rajasthani, Sindhi, Urdu
double tilde	ñ	Tagalog
double underscore	g h	Transliterated: Braj, Hindi, Maithili, Mewari, Nepali, Pahari, Rajasthani
grave accent	à è ì ñ ò ù š	Afrikaans, Catalan, French, Italian, Portuguese, Tagalog, Vietnamese, Wendic Transliterated: Khmer, Yiddish
háček	ǎ č ě ě ĝ ħ ň ő ř š ť ů ž	Czech, Latvian, Lithuanian, Serbo-Croatian (Latin), Slovak, Slovene, Wendic Transliterated: Amharic, Arabic, Armenian, Church Slavic, Georgian, Macedonian, Serbo-Croatian (Cyrillic), Sinhalese, Thai

(Continued on next page)

Table B1 (Continued)

Name	Examples	Languages
half circle below (upadhmaniya)	ḥ	Transliterated: Sanskrit, Prakrit
high comma, centered	ġ	Latvian
high comma, off center	d' g' k' l' t'	Czech, Navaho, Slovak, Wendic, Serbo-Croatian (Latin), Slovene
left hook	ą ɸ Ģ ǰ ɳ ɹ š ʃ ɛ ʝ ɸ	Latvian, Romanian, Polish
ligature	îă îê îô îu țș Țh ôț ȣș pș îê îq	Transliterated: Belorussian, Bulgarian, Church Slavic, Russian, Ukrainian
low rising tone mark	ă ẽ ư	Vietnamese
macron	ā æ ē ī k̄ ī n̄ ō œ p̄ r̄ s̄ ū	Anglo-Saxon, Hawaiian, Latvian, Lithuanian Transliterated: Amharic, Arabic, Armenian, Assamese, Bengali, Braj, Burmese, Church Slavic, Dogri, Georgian, Greek, Gujarati, Hindi, Japanese, Kannada, Khmer, Konkani, Korean, Lao, Lahnda, Maithili, Malayalam, Manipuri, Marathi, Mewari, Nepali, Oriya, Pahari, Panjabi, Persian, Prakrit, Pushto, Rajasthani, Russian, Sanskrit, Sindhi, Sinhalese, Tamil, Telugu, Thai, Tibetan, Urdu
right cedilla	ç	Transliterated: Thai
right hook	ą ɸ Ģ ǰ ɳ ɹ ʝ	Anglo-Saxon, Lithuanian, Navaho, Polish, Vietnamese Transliterated: Church Slavic, Lao
tilde	ã ẽ ī ñ õ q̃ s̃ ũ	Estonian, Navaho, Portuguese, Spanish, Vietnamese Transliterated: Amharic, Assamese, Bengali, Braj, Dogri, Gujarati, Hindi, Kannada, Khmer, Konkani, Lahnda, Lithuanian, Maithili, Malayalam, Manipuri, Marathi, Mewari, Nepali, Oriya, Pahari, Panjabi, Prakrit, Rajasthani, Sanskrit, Sindhi, Sinhalese, Tamil, Telugu, Tibetan
umlaut (diaeresis)	ä ë ğ ï j̈ ö ü v̈ ÿ	Afrikaans, Albanian, Catalan, Dutch, Estonian, French, German, Hungarian, Icelandic, Norwegian, Portuguese, Slovak, Spanish, Swedish, Turkish Transliterated: Chinese, Russian, Sindhi, Ukrainian

(Continued on next page)

Table B1 (Continued)

Name	Examples	Languages
underscore	<u>k</u> <u>l</u> <u>n</u> <u>r</u> <u>s</u> <u>t</u> <u>z</u> <u>g</u> <u>h</u> <u>d</u>	Transliterated: Assamese, Bengali, Braj, Dogri, Greek, Hindi, Kannada, Konkani, Lahnda, Maithili, Malayalam, Manipuri, Mewari, Nepali, Pahari, Panjabi, Persian, Prakrit, Pushto, Rajasthani, Sanskrit, Sindhi, Tamil, Telugu, Urdu

Table B2
Special Characters

Name	Examples	Languages
alif	’	Indonesian, Turkish Transliterated: Amharic, Arabic, Assamese, Bengali, Braj, Burmese, Dogri, Georgian, Gujarati, Hebrew, Hindi, Japanese, Khmer, Korean, Lao, Maithili, Malayalam, Manipuri, Marathi, Mewari, Nepali, Oriya, Pahari, Persian, Prakrit, Rajasthani, Sanskrit, Sinhalese, Tamil, Telugu, Tibetan, Urdu, Yiddish
‘ayn	‘	Navaho Transliterated: Amharic, Arabic, Armenian, Burmese, Chinese, Georgian, Hebrew, Khmer, Korean, Lao, Persian, Pushto, Sindhi, Thai, Urdu, Yiddish
degree sign	°	Transliterated: Khmer
dotless i	ı (I)	Turkish (Modern)
eth	ð (Ð)	Anglo-Saxon, Faroese, Icelandic (Modern)
hook o	ơ (Ơ)	Vietnamese
hook u	ư (Ư)	Vietnamese Transliterated: Thai
ligature æ	æ (Æ)	Anglo-Saxon, Danish, Faroese, Icelandic (Modern), Navaho, Norwegian Transliterated: Lao, Thai

(Continued on next page)

APPENDIX

Table B2 (Continued)

Name	Examples	Languages
ligature oe	œ (Œ)	Anglo-Saxon, French Transliterated: Lao, Thai
miâgkiî znak	'	Transliterated: Arabic, Belorussian, Burmese, Church Slavic, Hebrew, Khmer, Persian, Pushto, Russian, Tibetan, Ukrainian, Yiddish
middle dot	•	Catalan Transliterated: Khmer
slash d	đ (Đ)	Serbo-Croatian (Latin), Vietnamese Transliterated: Macedonian, Serbo-Croatian (Cyrillic)
slash l	ł (Ł)	Navaho, Polish, Wendic
slash o	ø (Ø)	Danish, Faroese, Norwegian
superscript two	²	Indonesian
thorn	þ (Þ)	Anglo-Saxon, Icelandic (Modern)
tvěrdyî znak	"	Transliterated: Burmese, Church Slavic, Russian, Ukrainian

NOTE: The superscript two,², is shown here for reference purposes only. It does not appear in the extended Latin alphabet coded character set as described in this standard.