First draft version of the
European Ordering Rules
Date:1998.04.30

Marc Wilhelm Küster

Table of Contents

Introduction

This European Standard is committed to facilitating crossborder communications and data exchange for the people of the European Community and the European Free Trade Area and to ensuring that European cultural requirements are safeguarded in the increasingly interconnected world of today.

This European Standard deals with rules for multilingual European ordering – the rules that come into effect if data from different languages must be brought into a sequence that makes it easy for users to find information.

Sorting is primarily a service for users to facilitate their access to information by presenting it in a structured way. It may contain subdividing information by subject matters, including several registers in a book, splitting a phone book into several sections, one for each town that falls into its purview, having multiple indices in a library... *Ordering* – the arrangement of information in alphabetical sequence – is in most circumstances an integral part of this procedure.

A standard can unfortunately be rather technical and difficult to read as it must present things in such a manner that an implementor can find unequivocal and complete guidelines. For anyone seriously enganged in ordering, be it for implementing it in a computer application or making library catalogues or any other field of action, it is unavoidable to deal with these technicalities. However, a beginner's introduction in plain English into this European Standard and implementations of this ordering standard in various formats are available on the World Wide Web via http://www.stri.is/TC304/EOR/[1] [2]

---

[1] ⟨NOTE⟩ Footnotes in ⟨NOTE⟩...⟨/NOTE⟩ pairs are not intended for inclusion into the standard, but should explain certain decisions or ommissions on the draft level.⟨/NOTE⟩

[2] ⟨NOTE⟩ The decision to place technical parts on this Web page has been taken for various reasons, the most prominent of which are:
 – Technical demands vary and no standardized or proprietary form of implementing the EOR should be discriminated against
 – Implementations tend to be error prone. On a Web site these errors can be readily corrected
 – Implementations are not part the ordering principles as such and tend to obfuscate them ⟨/NOTE⟩

1 Scope

This European Standard specifies the sequence to be established by alphabetical ordering of multilingual data composed of characters comprised in the *Multilingual European Subset Number 2* of *ISO/IEC 10646–1:1993* or subsets thereof.

NOTE The *Multilingual European Subset Number 2* is usually termed MES–2 and was formerly known as the *Minimal European Subset* (MES). Cf. *ENV 1923:1995*, currently under revision. MES–2 basically covers the Latin, Greek and Cyrillic letters needed in European data interchange and commerce and a variety of symbols

The ordering rules given here are only intended for data in more than one European language. They are not meant to replace or influence existing ordering rules for any single language.

The main part of this European Standard specifies letter-by-letter ordering of character strings. Normative Annex A deals with word-by-word ordering as a special form of ordering with multiple keys. Informative Annex B explains the use of further ordering criteria.

Informative Annex C presents a widely used alternative to the main part, namely the amalgamation of several scripts in one index via implicit transliteration.

2 Normative references

This European Standard incorporates by dated or undated reference provisions from other publications. These normative references are quoted at the appropriate places in the text, and the publications are listed hereafter.

All standards are subject to revision. Dated references do not refer to subsequent amendments of the publication in question. Undated references always refer to the latest edition.

ISO/IEC 10646–1:1993	Information Technology – Universal Multi-Octet Coded Character set (UCS). Including amendmends 1–4 and technical corrigendum 1
ISO/FDIS 12199:1997	Alphabetical ordering of multilingual terminological and lexicographical data represented in the Latin alphabet
ISO/IEC FCD 14651:1997	International String Ordering – Method for comparing Character Strings and Description of the Common Template Tailorable Ordering
ENV 1973:1995	Information Technology – European subsets of *ISO/IEC 10646–1:1993*
ENV 1973:XXXX[3]	InformationTechnology–Europeansubsetsof *ISO/IEC 10646–1:1993*[4]

---

[3] ⟨NOTE⟩ Must be replaced by the correct reference to the future MES–2⟨/NOTE⟩
[4] ⟨NOTE⟩ Must be replaced by the correct name of this forthcoming document⟨/NOTE⟩

3 Notation

Characters are referenced as U+XXXX where *X* stands for any hexadecimal value <= F and refers to the value of that character in *ISO/IEC 10646–1:1993*/ Unicode. This convention shall be used throughout this European Standard

4 Definitions

For the purpose of the European Standard, the following definitions apply:

4.1
character
unit of information used for the organization, control, or representation of textual data

NOTE For the purpose of this European Standard always a member of the *Multilingual European Subset No 2.*

4.2
digit
one of the characters `0123456789` (U+0030–U+0039)

4.3
letter
character used to represent (either alone or in combination) the sounds of a natural language in writing.[5] Here equivalent to all characters of the *Multilingual European Subset No 2*[6] with the property `LETTER` or `LIGATURE`

4.4
basic letter
character that is a member of the following list of letters:
a A b B c C d D e E f F g G h H i I j J k K l L m M n N o O p P q Q r R s S t T u U v V w W x X y Y z Z þ Þ (U+0061–007A, U+0041–005A, U+00FE, U+00DE)
α A β B γ Γ δ Δ ε E ζ Z η H θ Θ ι I κ K λ Λ μ M ν N ξ Ξ o O π Π ρ P σ Σ τ T υ Y φ Φ χ X ψ Ψ ω Ω (U+03B1–03C9, U+0391–03A9)
а А б Б в В г Г д Д е Е є Є ж Ж ѕ Ѕ з З и И й Й і I ј Ј к К л Л љ Љ м М н Н њ Њ о О п П р Р с С т Т ћ Ћ у У ф Ф х Х ц Ц ч Ч ш Ш щ Щ ъ Ъ ы Ы г Ґ э Э ю Ю я Я [7] (U+0430–044F, U+0410–042F)

4.5
diacritical mark
any of a number of recurring graphical strutures placed over, under, or through a basic letter which does not modify the shape of the basic letter itself and which in combination with that basic letter is a valid letter. These structures modify meaning or pronunciation or some other

---

[5] ⟨NOTE⟩ This definition has been thus termed to exclude punctuation marks from being interpreted as letters⟨/NOTE⟩

[6] ⟨NOTE⟩ As MES–2 is not yet available, all tables in this draft are based on the *Minimal European Subset* (ENV 1973:1995)⟨/NOTE⟩

[7] ⟨NOTE⟩ This list is based on a list transmitted by John Clews⟨/NOTE⟩

feature of the basic letter. The diacritical marks which are relevant to this European Standard are listed in Table 1

4.6
letter with diacritical marks
letter which is constructed as the combination between a basic letter and a diacritical mark

4.7
equivalent letter form
character created by joining two or more distinct basic letters or two or more letter with diacritical mark or any combination of these. The equivalent letter forms which are relevant to this European Standard are listed in Table 3

4.8
modified letter
letter that is neither a basic letter nor an equivalent letter form nor a letter with diacritical mark. The modified letters which are relevant to this European Standard are listed in Table 4

4.9
uppercase letter
one of the basic letters

  ABCDEFGHIJKLMNOPQRSTUVWXYZÞ
  ΑΒΓΔΕΖΗΘΙΚΛΜΝΞΟΠΡΣΤΥΦΧΨΩ
  АБВГДЕЄЖЅЗИЙІЈКЛЉМНЊОПРСТЋУФХЦЧШЩЪЬЫГЭЮЯ

or a letter with diacritical mark formed with one of these basic letters or a modified letter qualified as CAPITAL in Table 4

4.10
lowercase letter
letter which is not an uppercase letter

4.11
special character
character that is neither a letter nor a digit

4.12
character string
series of characters in sequence. Hereafter often referenced as string

4.13
ordering
act of bringing character strings into a well-defined sequence according to a string comparison specification

4.14
space character
one of the characters SPACE (U+0020) and NO-BREAK SPACE (U+00A0)

5 Preparatory procedures

5.1 Motivation

Most ordering tasks require more than simply ordering of strings. In a telephone directory, for example, one might want to order by names first, followed by addresses and phone numbers, recurring to addresses only when ordering by names fails to establish a unique sequence and to phone numbers only if both names and addresses are equal.

Each of these units is called a *key* and the approach the *multiple ordering key* approach.

5.2 Rules

More rigorously expressed, the multiple ordering key approach implies the preprocessing of the data in the following steps, any or all of which may be omitted, especially in the case of a single ordering key:
- subdivision of data into multiple ordering keys through the introduction of a higher level protocol
- establishing a hierarchy between these keys
- extracting the keys from the data
- subjecting the keys to some form of normalization

NOTE This normalization might include, but is not limited to: changing upper-case letters to lower case where it is considered appropriate (e. g. in the case of sentence initial capitals or capitals for emphasis), lemmatization (especially for inflected languages), expansion of abbreviations, insertion of leading zeroes in numbers or reduction of blanks between words to one throughout the data. It can also be left out entirely

- Starting with the keys highest in the hierarchy equivalent keys which were thus obtained are compared with the aid of the ordering rules as established in this European Standard. As soon as a unique sequence is established, further keys are ignored.

5.3 Further preprocessing

Further preprocessing of some kind may or may not be necessary, but is not within the scope of this European Standard.

This European Standard assumes that the user has already performed these preparatory procedures which are left entirely at his or her discretion and are thus out of its scope. It is concerned exclusively with the ordering of strings which belong to one key and which have undergone those preparatory procedures.

6 The multilevel ordering procedure

6.1 General principles

This European Standards defines a multilevel ordering procedure which means that the input strings are first compared on the *primary ordering level.* Only when the procedure described for this level fails to establish a unique and determined sequence for the strings the different parts of the *second ordering level* are taken into consideration. If this likewise fails to produce a unique sequence the *tertiary ordering level* is invoked. If this also cannot establish a unique sequence two strings are regarded as equivalent.

Each level compares two strings in the following manner: The first non-ignored characters are compared. If the ordering rules for that level specify a unique and determined sequence for these characters then this determines the sequence of the strings. If not, the second non-ignored characters are compared, and so forth until one of the following conditions is met. In the case that more than one of the conditions are true, only the first one which is fulfilled is applicable:

1.) the ordering rules for that level define a unique sequence for the two non-ignored characters which is then also the ordering sequence for the strings
2.) one of the strings has no more non-ignored characters whereas the other has. Then the string without more characters precedes the other one
3.) both strings have no more non-ignored characters. Then the next ordering level, if existing, is invoked. If there are no more levels, the two strings are deemed equivalent

6.2 Assumptions and aims

This European Standard acts from a number of assumptions:
– access to information must be facilitated as much as possible;
– the user is not assumed to know details of *ISO/IEC 10646–1:1993;*
– the rules are derived from standardized rules and common practice in a large number of European languages without giving preference to the rules of any language or languages in particular.

NOTE These assumptions motivate a set of principles which underly these European Ordering Rules and help to clarify the decisions taken.
– modified letters are ordered according to their visual appearance, not according to their pronunciation or meaning unless user-expectation demands something else;
– Forms which the user perceives as more basic should precede special or combined ones.

6.3 Axioms (valid throughout)

6.3.1 Ordering by script

Digits precede letters. Letters are ordered by scripts, putting Latin letters before Greek ones before Cyrillic ones.

6.3.2 Equivalent letter forms

Equivalent letter forms are decomposed into the letters of which they are formed as specified in Table 3

7 Primary ordering level

7.1 Validity

All of the following rules are valid for the primary ordering level only.

7.2 Equivalent or ignored characters

7.2.1 Upper- and lowercase letters

Upper- and lowercase forms of the same letter are treated as equivalent

7.2.2 Modified letters

Modified letters are treated as equivalent to one or more basic letters as specified in Table 4

7.2.3 Letters with diacritical marks

Letters with diacritical marks are treated as equivalent to their corresponding basic letters

7.2.4 Special characters

Special characters are ignored

7.3 Ordering sequences

7.3.1 Digits

Digits are to be ordered in the following sequence:
    0123456789

7.3.2 Latin script

Basic Latin letters are to be ordered in the following sequence:
    ABCDEFGHIJKLMNOPQRSTUVWXYZÞ

7.3.3 Greek script

Basic Greek letters are to be ordered in the following sequence:
    ΑΒΓΔΕΖΗΘΙΚΛΜΝΞΟΠΡΣΤΥΦΧΨΩ

NOTE 1 This is the sequence of the relevant Greek Ordering Standard ELOT **XXX**

7.3.4 Cyrillic script

Basic Cyrillic letters are to be ordered in the following sequence:
АБВГД ЕЄЖЅЗИЙІЈКЛЉЬМ НЊОПРСТЋУФХЦЧШЩЪЫГЭЮЯ

NOTE 2 The sequence was obtained by accepting the Old Slavonic ordering and applying it to the Cyrillic characters of MES–2. For the Russian subset this results in the expected ordering sequence for that language

8 Secondary ordering level

8.1 No unique sequence after the primary ordering level

If the primary ordering level does not result in a unique sequence, the secondary ordering level is invoked. It is distinguished from the primary ordering level by no longer treating letters with diacritical marks and modified letters as equivalent to basic letters.

The secondary ordering level is divided into two parts: modified letters and diacritical marks. If the treatment of modified letters alone results in a unique sequence diacritical marks are to be ignored.

8.2 Equivalent or ignored characters

8.2.1 Upper- and lowercase letters

Upper- and lowercase forms of the same letter are treated as equivalent.

8.2.2 Special characters

Special characters are ignored

8.3 Ordering sequences

8.3.1 Modified letters

Modified letters are to be ordered after their corresponding basic letter. In the case of multiple modified letters with the same basic letter they are to be ordered in the sequence indicated in Table 4.

8.3.2 Letters with diacritical marks

Letters with diacritical marks which have only one diacritical mark are to be ordered with respect to their diacritical mark in the sequence indicated in Table 1. For letters with more than one diacritical mark, the diacritical mark shall be considered in the following order:

Inside the character before outside; above the character before below; working from top to bottom, then from left to right.[8] In practice, this results for MES–2 in the sequence indicated in Table 2.

9 Tertiary ordering level

9.1 No unique sequence after the secondary ordering level

If the secondary ordering level also does not result in a unique sequence of strings, the third ordering level is invoked. It does no longer treat uppercase and lowercase letters as equivalent and takes special characters into account.

The tertiary level is divided into two parts. If the capitalization alone suffices to establish a unique sequence special characters are ignored.

9.2 Ordering sequences

9.2.1 Capitalization

Lowercase letters are ordered before the corresponding uppercase ones.[9]

9.2.2 Special characters

The special characters of the *Multilingual European Subset No 2* are subdivided into four categories whose repertoires are listed in the respective tables.
  1.) Special Characters: Punctuation marks and spacing diacritics (Table 5 and Table 6)
  2.) Special Characters: Characters for commerce   (Table 7)
  3.) Special Characters: Mathematical symbols   (Table 8)
  4.) Special Characters: Unqualified   (Table 9)
  Special characters belonging to the same category are to be ordered in the sequence given in the respective table. Special characters which belong to different categories are to be ordered in the sequence of their categories as indicated in the list above.

---

[8] ⟨NOTE⟩ This is a modified version of the clause of *ISO 12199* with which this European Standard should not conflict. It was changed to cater for letters such as å̊⟨/NOTE⟩

[9] ⟨NOTE⟩ This is a point of heated discussions for some, as the general principle »basic form before special form« gives no clear indications. Historically, capital letters are more ancient, there having been no small letters in antiquity. Nowadays, most people would intuitively regard lowercase letters as the more basic form, especially in languages which do not capitalize nouns.
A survey by Håvard Hjulstad shows that of the national ordering standards listed in the bibliography three (Canadia, Germany, Sweden) put lowercase before uppercase, one (Denmark) uppercase before lowercase and four (Great Britain, Finland, France and Norwegia) do not have a fixed rule ⟨/NOTE⟩

9.3 Equivalence

Two strings between which after the tertiary ordering level no unique sequence can be established are considered as being equivalent.

NOTE For further options to break the deadlock in certain circumstances please cf. the informative annex B: *Ordering by position and by style*

10 Tables

10.1 Diacritical marks

10.1.1 Diacritical marks

Table 1 – Diacritical marks[10]

Shape[11]	Corresponding diacritical mark in ISO/IEC 10646[12]	Alternative names[13]
	Diacritical marks for the Greek script	
᾿	U+1FBF PSILI	spiritus lenis
῾	U+1FFE DASIA	spiritus asper
`	U+1FEF VARIA	
´	U+1FFD OXIA	
�‿	U+0306 COMBINING BREVE	VRACHY
῀	U+0342 COMBINING GREEK PERISPOMENI	
΄	U+0384 TONOS	
ι	U+1FBE PROSGEGRAMMENI	iota adscriptum[14]
	U+0345 COMBINING GREEK YPOGEGRAMMENI	iota subscriptum[15]
¨	U+0308 COMBINING DIAERESIS	DIALYTICA
¯	U+0304 COMBINING MACRON	Greek macron, length
	Diacritical marks for Latin and Cyrillic scripts	
´	U+0301 COMBINING ACUTE ACCENT	
`	U+0300 COMBINING GRAVE ACCENT	
�‿	U+0306 COMBINING BREVE	
^	U+0302 COMBINING CIRCUMFLEX ACCENT	
ˇ	U+030C COMBINING CARON	
°	U+030A COMBINING RING ABOVE	
¨	U+0308 COMBINING DIAERESIS	umlaut, trema[16]
˝	U+030B COMBINING DOUBLE ACUTE	
~	U+0303 COMBINING TILDE	
·	U+0307 COMBINING DOT ABOVE	
¸	U+0327 COMBINING CEDILLA	
˛	U+0328 COMBINING OGONEK	
₋	U+0304 COMBINING MACRON	

---

[10] ⟨NOTE⟩ The formatting of these tables is as yet temporary and does not conform to ISO guidelines, especially with respect to the handling of notes. This will be changed as soon as the draft stabilizes. Only diacritical marks that are acutally needed for the MES–2 are listed here. The 'preceded-by-apostrophe' which occurrs only in LATIN SMALL LETTER N PRECEDED BY APOSTROPHE (U+0146, 'n) has been ignored as this letter is both flagged as being Afrikaans which is not an indigenous language of Europe and as being *not actually a single letter*⟨/NOTE⟩

[11] Shapes are informative only

[12] Position and name of the corresponding diacritical mark in *ISO/IEC 10646–1:1993*. If possible, combining diacritical marks are referenced. *Combining* is a character property in *ISO/IEC 10646–1:1993*. If no corresponding combining diacritical mark exists, the table lists non-combining variants. Diacritical marks are unified for Cyrillic and Latin but not for Greek and Latin. This reflects prevalent usage and user-expectations

[13] Names in lowercase letters are only an informative selection of some of the most common alternative names. Names in capitals are normative

[14] The prosgegrammeni / *iota adscriptum* is treated as a diacritical mark at the demand of the Greek standards organization ELOT

[15] Exists only in combination with α, η, ω as ᾳ, ῃ, ῳ

[16] Strictly speaking, umlaut and trema can be two typographically slightly differerent phenonmena, but the distinction is increasingly becoming obsolete

This form of presentation has been so chosen that an implementor of a conforming ordering application can, if she or he so chooses, unify the treatment of diacritical marks accross scripts without modifying the resulting sequence of strings.[17]

10.1.2 Multiple diacritical marks

Table 2 – Multiple diacritical marks

Shape	Corresponding diacritical mark in ISO/IEC 10646[18]
	Multiple diacritical marks for the Greek script
᾽	U+1FCD PSILI AND VARIA
[19]	– PSILI AND VARIA AND PROSGEGRAMMENI
῾	– PSILI AND VARIA AND YPOGEGRAMMENI
	U+1FCE PSILI AND OXIA
	– PSILI AND OXIA AND PROSGEGRAMMENI
῏	– PSILI AND OXIA AND YPOGEGRAMMENI
	U+1FCF PSILI AND PERISPOMENI
	– PSILI AND PERISPOMENI AND PROSGEGRAMMENI
	– PSILI AND PERISPOMENI AND YPOGEGRAMMENI
	– PSILI AND PROSGEGRAMMENI
῝	– PSILI AND YPOGEGRAMMENI
	U+1FDD DASIA AND VARIA
	– DASIA AND VARIA AND PROSGEGRAMMENI
῟	– DASIA AND VARIA AND YPOGEGRAMMENI
	U+1FDE DASIA AND OXIA
	– DASIA AND OXIA AND PROSGEGRAMMENI
῎	– DASIA AND OXIA AND YPOGEGRAMMENI
	U+1FDF DASIA AND PERISPOMENI
	– DASIA AND PERISPOMENI AND PROSGEGRAMMENI
	– DASIA AND PERISPOMENI AND YPOGEGRAMMENI
	– DASIA AND PROSGEGRAMMENI
	– DASIA AND YPOGEGRAMMENI
	– VARIA AND YPOGEGRAMMENI
	– OXIA AND YPOGEGRAMMENI
῭	– PERISPOMENI AND YPOGEGRAMMENI
΅	U+1FED DIALYTIKA AND VARIA
῁	U+1FEE DIALYTIKA AND OXIA
῍	U+1FC1 DIALYTIKA AND PERISPOMENI
	U+0385 DIALYTIKA AND TONOS
᾿	Multiple diacritical marks for the Latin script
῀	– RING ABOVE AND ACUTE
	– DIAERESIS AND MACRON

---

[17] ⟨NOTE⟩ It does not modify the ordering sequence if the Greek and Latin sections are exchanged⟨/NOTE⟩

[18] Position and name in *ISO/IEC 10646–1:1993*. If no corresponding combining diacritical mark exists, the table lists non-combining variants. If these also do not exist, the table simply gives the names of the diacritical marks. Diacritical marks are not unified accross scripts unless this reflects prevalent usage and user-expectations

[19] This and several other combinations cannot reasonably be printed

10.2 Equivalent forms

Table 3 – Equivalent forms

Shape	Position and name in ISO/IEC 10646	Equivalent
æ	U+00E6 LATIN SMALL LETTER AE	ae
Æ	U+00C6 LATIN CAPITAL LETTER AE	Ae
ǽ	U+01FD LATIN SMALL LETTER AE WITH ACUTE	áe
Ǽ	U+01FC LATIN CAPITAL LETTER AE WITH ACUTE	Áe
dz	U+01F3 LATIN SMALL LETTER DZ	dz
DZ	U+01F1 LATIN CAPITAL LETTER DZ	DZ
dž	U+01C6 LATIN SMALL LETTER DZ WITH CARON	dž
DŽ	U+01C4 LATIN CAPITAL LETTER DZ WITH CARON	DŽ
fi	U+FB01 LATIN SMALL LIGATURE FI	fi
fl	U+FB02 LATIN SMALL LIGATURE FL	fl
ĳ	U+0133 LATIN SMALL LIGATURE IJ	ij
Ĳ	U+0132 LATIN CAPITAL LIGATURE IJ	IJ
lj	U+01C9 LATIN SMALL LETTER LJ	lj
LJ	U+01C7 LATIN CAPITAL LETTER LJ	LJ
nj	U+01CC LATIN SMALL LETTER NJ	nj
NJ	U+01CA LATIN CAPITAL LETTER NJ	NJ
œ	U+0153 LATIN SMALL LIGATURE OE	oe
Œ	U+0152 LATIN CAPITAL LIGATURE OE	Oe

10.3 Modified letters

Table 4 – Modified letters

Shape	Position and name of special letter in ISO/IEC 10646	Equiv. POL[20]
đ	U+0111 LATIN SMALL LETTER D WITH STROKE	d
Đ	U+0110 LATIN CAPITAL LETTER D WITH STROKE	D
ð	U+00F0 LATIN SMALL LETTER ETH	d
Ð	U+00D0 LATIN CAPITAL LETTER ETH	D
ǥ	U+01E5 LATIN SMALL LETTER G WITH STROKE	g
Ǥ	U+01E4 LATIN CAPITAL LETTER G WITH STROKE	G
ħ	U+0127 LATIN SMALL LETTER H WITH STROKE	h
Ħ	U+0126 LATIN CAPITAL LETTER H WITH STROKE	H
ı	U+0131 LATIN SMALL LETTER DOTLESS I	i
İ	U+0130 LATIN CAPITAL LETTER I WITH DOT ABOVE	I
	U+0192 LATIN SMALL LETTER F WITH HOOK	f
	U+0138 LATIN SMALL LETTER KRA	k
ł	U+0142 LATIN SMALL LETTER L WITH STROKE	l
Ł	U+0141 LATIN CAPITAL LETTER L WITH STROKE	L
ŀ	U+0140 LATIN SMALL LETTER L WITH MIDDLE DOT	l
Ŀ	U+013F LATIN CAPITAL LETTER L WITH MIDDLE DOT	L
ⁿ	U+207F SUPERSCRIPT LATIN SMALL LETTER N	n
'n	U+0149 LATIN SMALL LETTER N PRECEDED BY APOSTROPHE	n
	U+014B LATIN SMALL LETTER ENG	ng
	U+014A LATIN CAPITAL LETTER ENG	Ng
ø	U+00F8 LATIN SMALL LETTER O WITH STROKE	o
Ø	U+00D8 LATIN CAPITAL LETTER O WITH STROKE	O
	U+027C LATIN SMALL LETTER R WITH LONG LEG	r

---

[20] Equivalant on Primary Ordering Level

ſ	U+017F LATIN SMALL LETTER LONG S	s
ß	U+00DF LATIN SMALL LETTER SHARP S	ss
ŧ	U+0167 LATIN SMALL LETTER T WITH STROKE	t
Ŧ	U+0166 LATIN CAPITAL LETTER T WITH STROKE	T
ʒ	U+0292 LATIN SMALL LETTER EZH	z
Ʒ	U+01B7 LATIN CAPITAL LETTER EZH	Z
ǯ	U+01EF LATIN SMALL LETTER EZH WITH CARON	z
Ǯ	U+01EE LATIN CAPITAL LETTER EZH WITH CARON	Z
ς	U+03C2 FINAL SIGMA	σ
ґ	U+0491 CYRILLIC SMALL LETTER GHE WITH UPTURN	г
Ґ	U+0490 CYRILLIC CAPITAL LETTER GHE WITH UPTURN	Г

10.4 Special characters

10.4.1 Special characters: Punctuation marks and spacing diacritics

Table 5 – Special characters: Punctuation marks

Shape   Position and name of special character in ISO/IEC 10646

Shape	Position and name of special character in ISO/IEC 10646
	U+0020 SPACE
!	U+0021 EXCLAMATION MARK
"	U+0022 QUOTATION MARK
'	U+0027 APOSTROPHE
(	U+0028 LEFT PARENTHESIS
)	U+0029 RIGHT PARENTHESIS
,	U+002C COMMA
−	U+002D HYPHEN-MINUS
.	U+002E FULL STOP
/	U+002F SOLIDUS
:	U+003A COLON
;	U+003B SEMICOLON
?	U+003F QUESTION MARK
[	U+005B LEFT SQUARE BRACKET
\	U+005C REVERSE SOLIDUS
]	U+005D RIGHT SQUARE BRACKET
{	U+007B LEFT CURLY BRACKET
}	U+007D RIGHT CURLY BRACKET
	U+00A0 NO-BREAK SPACE
¡	U+00A1 INVERTED EXCLAMATION MARK
«	U+00AB LEFT-POINTING DOUBLE ANGLE QUOTATION MARK
	U+00AD SOFT HYPHEN
»	U+00BB RIGHT-POINTING DOUBLE ANGLE QUOTATION MARK
¿	U+00BF INVERTED QUESTION MARK
;	U+037E GREEK QUESTION MARK
·	U+0387 GREEK ANO TELEIA
‐	U+2010 HYPHEN
–	U+2013 EN DASH
—	U+2014 EM DASH
'	U+2018 LEFT SINGLE QUOTATION MARK
'	U+2019 RIGHT SINGLE QUOTATION MARK
‚	U+201A SINGLE LOW-9 QUOTATION MARK
	U+201B SINGLE HIGH-REVERSED-9 QUOTATION MARK
"	U+201C LEFT DOUBLE QUOTATION MARK
"	U+201D RIGHT DOUBLE QUOTATION MARK
„	U+201E DOUBLE LOW-9 QUOTATION MARK
‹	U+2039 SINGLE LEFT-POINTING ANGLE QUOTATION MARK

›   U+203A SINGLE RIGHT–POINTING ANGLE QUOTATION MARK
      U+203C DOUBLE EXCLAMATION MARK

Table 6 – Special characters: Spacing diacritics[21]

Shape	Position and name of special character in ISO/IEC 10646
᾿	U+1FBF GREEK PSILI
῍	U+1FCD GREEK PSILI AND VARIA
῎	U+1FCE GREEK PSILI AND OXIA
῏	U+1FCF GREEK PSILI AND PERISPOMENI
῾	U+1FFE GREEK DASIA
῝	U+1FDD GREEK DASIA AND VARIA
῞	U+1FDE GREEK DASIA AND OXIA
῟	U+1FDF GREEK DASIA AND PERISPOMENI
`	U+1FEF GREEK VARIA
´	U+1FFD GREEK OXIA
῀	U+1FC0 GREEK PERISPOMENI
'	U+0384 GREEK TONOS
ι	U+1FBE GREEK PROSGEGRAMMENI
ͺ	U+037A GREEK YPOGEGRAMMENI
´	U+00B4 ACUTE ACCENT
`	U+0060 GRAVE ACCENT
˘	U+02D8 BREVE
᾿	U+1FBD GREEK KORONIS
^	U+005E CIRCUMFLEX ACCENT
ˇ	U+02C7 CARON
°	U+02DA RING ABOVE
¨	U+00A8 DIAERESIS
῭	U+1FED GREEK DIALYTIKA AND VARIA
΅	U+1FEE GREEK DIALYTIKA AND OXIA
῁	U+1FC1 GREEK DIALYTIKA AND PERISPOMENI
΅	U+0385 GREEK DIALYTIKA AND TONOS
˝	U+02DD DOUBLE ACUTE ACCENT
~	U+007E TILDE
˜	U+02DC SMALL TILDE
˙	U+02D9 DOT ABOVE
¸	U+00B8 CEDILLA
˛	U+02DB OGONEK
¯	U+00AF MACRON

10.4.2 Special characters: Characters for commerce

Table 7 – Special characters: Characters for commerce

Shape	Position and name of special character in ISO/IEC 10646
#	U+0023 NUMBER SIGN
$	U+0024 DOLLAR SIGN
%	U+0025 PERCENT SIGN
&	U+0026 AMPERSAND
@	U+0040 COMMERCIAL AT
¢	U+00A2 CENT SIGN
£	U+00A3 POUND SIGN
¤	U+00A4 CURRENCY SIGN

---

[21] ⟨NOTE⟩ In contrast to the other tables of special characters the spacing diacritics are not ordered according in the sequence of *ISO/IEC 10646–1* but in the same way diacritical marks are treated on the secondary ordering level. This is more in line with user expectations⟨/NOTE⟩

¥	U+00A5 YEN SIGN
©	U+00A9 COPYRIGHT SIGN
	U+00AE REGISTERED SIGN
‰	U+2030 PER MILLE SIGN
	U+20A3 FRENCH FRANC SIGN
	U+20A4 LIRA SIGN
	U+20A7 PESETA SIGN
	U+2105 CARE OF
	U+2116 NUMERO SIGN
	U+2122 TRADE MARK SIGN
e	U+212E ESTIMATED SYMBOL

10.4.3 Special characters: Mathematical symbols

Table 8 – Special characters: Mathematical symbols

Shape	Position and name of special character in ISO/IEC 10646
+	U+002B PLUS SIGN
<	U+003C LESS-THAN SIGN
=	U+003D EQUALS SIGN
>	U+003E GREATER-THAN SIGN
¬	U+00AC NOT SIGN
°	U+00B0 DEGREE SIGN
±	U+00B1 PLUS-MINUS SIGN
^{2}	U+00B2 SUPERSCRIPT TWO
^{3}	U+00B3 SUPERSCRIPT THREE
μ	U+00B5 MICRO SIGN
·	U+00B7 MIDDLE DOT
^{1}	U+00B9 SUPERSCRIPT ONE
	U+00BC VULGAR FRACTION ONE QUARTER
	U+00BD VULGAR FRACTION ONE HALF
	U+00BE VULGAR FRACTION THREE QUARTERS
×	U+00D7 MULTIPLICATION SIGN
÷	U+00F7 DIVISION SIGN
	U+0374 GREEK NUMERAL SIGN
	U+0375 GREEK LOWER NUMERAL SIGN
′	U+2032 PRIME
″	U+2033 DOUBLE PRIME
⁄	U+2044 FRACTION SLASH
Ω	U+2126 OHM SIGN
	U+215B VULGAR FRACTION ONE EIGHTH
	U+215C VULGAR FRACTION THREE EIGHTHS
	U+215D VULGAR FRACTION FIVE EIGHTHS
	U+215E VULGAR FRACTION SEVEN EIGHTHS
∂	U+2202 PARTIAL DIFFERENTIAL
Δ	U+2206 INCREMENT
∏	U+220F N-ARY PRODUCT
∑	U+2211 N-ARY SUMMATION
−	U+2212 MINUS SIGN
•	U+2219 BULLET OPERATOR
√	U+221A SQUARE ROOT
∞	U+221E INFINITY
∩	U+2229 INTERSECTION
∫	U+222B INTEGRAL
≈	U+2248 ALMOST EQUAL TO
≠	U+2260 NOT EQUAL TO
≡	U+2261 IDENTICAL TO

16

≤	U+2264 LESS-THAN OR EQUAL TO
≥	U+2265 GREATER-THAN OR EQUAL TO
	U+2310 REVERSED NOT SIGN
⌠	U+2320 TOP HALF INTEGRAL
⌡	U+2321 BOTTOM HALF INTEGRAL

10.4.4 Special characters: unqualified

Table 9 – Special characters: unqualified

Shape	Position and name of special character in ISO/IEC 10646
*	U+002A ASTERISK
_	U+005F LOW LINE
		U+007C VERTICAL LINE
	U+00A6 BROKEN BAR
§	U+00A7 SECTION SIGN
ª	U+00AA FEMININE ORDINAL INDICATOR
¶	U+00B6 PILCROW SIGN
º	U+00BA MASCULINE ORDINAL INDICATOR
—	U+2015 HORIZONTAL BAR
	U+2017 DOUBLE LOW LINE
†	U+2020 DAGGER
	U+2021 DOUBLE DAGGER
	U+2022 BULLET
	U+2023 TRIANGULAR BULLET
...	U+2026 HORIZONTAL ELLIPSIS
	U+2028 LINE SEPARATOR
	U+2029 PARAGRAPH SEPARATOR
‾	U+203E OVERLINE
	U+2043 HYPHEN BULLET
	U+2190 LEFTWARDS ARROW
	U+2191 UPWARDS ARROW
	U+2192 RIGHTWARDS ARROW
	U+2193 DOWNWARDS ARROW
	U+2194 LEFT RIGHT ARROW
	U+2195 UP DOWN ARROW
	U+21A8 UP DOWN ARROW WITH BASE
	U+221F RIGHT ANGLE
	U+2302 HOUSE
	U+2500 BOX DRAWINGS LIGHT HORIZONTAL
	U+2502 BOX DRAWINGS LIGHT VERTICAL
	U+250C BOX DRAWINGS LIGHT DOWN AND RIGHT
	U+2510 BOX DRAWINGS LIGHT DOWN AND LEFT
	U+2514 BOX DRAWINGS LIGHT UP AND RIGHT
	U+2518 BOX DRAWINGS LIGHT UP AND LEFT
	U+251C BOX DRAWINGS LIGHT VERTICAL AND RIGHT
	U+2524 BOX DRAWINGS LIGHT VERTICAL AND LEFT
	U+252C BOX DRAWINGS LIGHT DOWN AND HORIZONTAL
	U+2534 BOX DRAWINGS LIGHT UP AND HORIZONTAL
	U+253C BOX DRAWINGS LIGHT VERTICAL AND HORIZONTAL
	U+2550 BOX DRAWINGS DOUBLE HORIZONTAL
	U+2551 BOX DRAWINGS DOUBLE VERTICAL
	U+2552 BOX DRAWINGS DOWN SINGLE AND RIGHT DOUBLE
	U+2553 BOX DRAWINGS DOWN DOUBLE AND RIGHT SINGLE
	U+2554 BOX DRAWINGS DOUBLE DOWN AND RIGHT
	U+2555 BOX DRAWINGS DOWN SINGLE AND LEFT DOUBLE
	U+2556 BOX DRAWINGS DOWN DOUBLE AND LEFT SINGLE

```
    U+2557 BOX DRAWINGS DOUBLE DOWN AND LEFT
    U+2558 BOX DRAWINGS UP SINGLE AND RIGHT DOUBLE
    U+2559 BOX DRAWINGS UP DOUBLE AND RIGHT SINGLE
    U+255A BOX DRAWINGS DOUBLE UP AND RIGHT
    U+255B BOX DRAWINGS UP SINGLE AND LEFT DOUBLE
    U+255C BOX DRAWINGS UP DOUBLE AND LEFT SINGLE
    U+255D BOX DRAWINGS DOUBLE UP AND LEFT
    U+255E BOX DRAWINGS VERTICAL SINGLE AND RIGHT DOUBLE
    U+255F BOX DRAWINGS VERTICAL DOUBLE AND RIGHT SINGLE
    U+2560 BOX DRAWINGS DOUBLE VERTICAL AND RIGHT
    U+2561 BOX DRAWINGS VERTICAL SINGLE AND LEFT DOUBLE
    U+2562 BOX DRAWINGS VERTICAL DOUBLE AND LEFT SINGLE
    U+2563 BOX DRAWINGS DOUBLE VERTICAL AND LEFT
    U+2564 BOX DRAWINGS DOWN SINGLE AND HORIZONTAL DOUBLE
    U+2565 BOX DRAWINGS DOWN DOUBLE AND HORIZONTAL SINGLE
    U+2566 BOX DRAWINGS DOUBLE DOWN AND HORIZONTAL
    U+2567 BOX DRAWINGS UP SINGLE AND HORIZONTAL DOUBLE
    U+2568 BOX DRAWINGS UP DOUBLE AND HORIZONTAL SINGLE
    U+2569 BOX DRAWINGS DOUBLE UP AND HORIZONTAL
    U+256A BOX DRAWINGS VERTICAL SINGLE AND HORIZONTAL DOUBLE
    U+256B BOX DRAWINGS VERTICAL DOUBLE AND HORIZONTAL SINGLE
    U+256C BOX DRAWINGS DOUBLE VERTICAL AND HORIZONTAL
    U+2580 UPPER HALF BLOCK
    U+2584 LOWER HALF BLOCK
    U+2588 FULL BLOCK
    U+258C LEFT HALF BLOCK
    U+2590 RIGHT HALF BLOCK
    U+2591 LIGHT SHADE
    U+2592 MEDIUM SHADE
    U+2593 DARK SHADE
    U+25A0 BLACK SQUARE
    U+25AC BLACK RECTANGLE
    U+25B2 BLACK UP-POINTING TRIANGLE
    U+25BA BLACK RIGHT-POINTING POINTER
    U+25BC BLACK DOWN-POINTING TRIANGLE
    U+25C4 BLACK LEFT-POINTING POINTER
◊   U+25CA LOZENGE
    U+25CB WHITE CIRCLE
    U+25D8 INVERSE BULLET
    U+25D9 INVERSE WHITE CIRCLE
    U+263A WHITE SMILING FACE
    U+263B BLACK SMILING FACE
    U+263C WHITE SUN WITH RAYS
    U+2640 FEMALE SIGN
    U+2642 MALE SIGN
♠   U+2660 BLACK SPADE SUIT
♣   U+2663 BLACK CLUB SUIT
♥   U+2665 BLACK HEART SUIT
♦   U+2666 BLACK DIAMOND SUIT
    U+266A EIGHTH NOTE
    U+266B BEAMED EIGHTH NOTES
```

Annex A (normative): Word-by-word ordering

A.1 Modified terminology

For the purpose of this appendix a **special character** shall be a character that is neither a letter nor a digit nor a diacritical mark nor a space character

A.2 Principles

Word-by-word ordering is a frequently used alternative to letter-by-letter ordering. It is a special case of multiple-key ordering which treats space characters as key seperators. The maximal string is thus a set of characters enclosed by space characters.

NOTE The string can well be smaller if further keys so demand

The sets of strings thus obtained are ordered following the European Ordering Rules as specified in the main part of this European Standard.

A.3 Example of Word-by-word vs. letter-by-letter ordering

Letter-by-letter odering	Word-by-word ordering
in-	in-
inability	in absentia
in absentia	in extenso
inadvisable	in medias res
in extenso	in memoriam
in medias res	inability
in memoriam	inadvisable

A.4 Simplified word-by-word ordering

If the text to be ordered word by word contains few modified letters, letters with diacritical marks, or special characters, the following method will in most cases produce the same result as the method that is specified above.

In the *ordering by script* section (6.3.1) spacing characters precede digits and letters. The space character is then removed from the table of special characters (Table 5 (10.4.1)). The other ordering rules remain unchanged.

Annex B (informative): Ordering by position and by style[22]

B.1 Motivation

In some cases it is desirable to differentiate further on the tertiary ordering level, e. g. in the case where definitions and ordinary usage of a word is categorized solely by the application of some form of tagging which in print usually takes the form of a formatting style. Another example might be the distinction between loan words and native words in such a manner.

This formatting can be expressed by changing the position to the baseline, e. g. in mathematical or chemical formulae, or by highlightening it with certain typgraphic features, e. g. italic typeface, that serves to indicate some property of the word.

B.2 Recommended rules

This European Standard recommends that, if the implementor deems it necessary to make this differentiation, she or he modify (9.2.1) (Capitilization) on the tertiary ordering level in the following manner:

Letters are to be arranged in the sequence indicated in this list:
1.) lowercase letter on baseline
2.) uppercase letter on baseline
3.) lowercase letter above baseline
4.) uppercase letter above baseline
5.) lowercase letter below baseline
6.) uppercase letter below baseline

If this does not result in a unique sequence, typographic styles are to be taken into consideration in the sequence listed:
1.) roman abcde
2.) boldface **abcde**
3.) italic *abcde*
4.) boldface italic ***abcde***
5.) others

Annex C (informative): Mixed-script ordering with one predominant script (»encyclopedia mode«)

C.1 Motivation

Many publications – often of the encyclopedia type – handle scripts differently from this European Standard, especially if they cover predominantly one script with a few entries from other scripts interspersed. They implicitly transliterate strings from other scripts into the predominant one and order according to the rules for that script. For printing the strings are then rendered in their original form. This has the advantage for the user to find related articles e. g. on λόγος and logic near to each other.

---

[22] ⟨NOTE⟩ This section is included for compatibility with *ISO FDIS 12199*⟨/NOTE⟩

NOTE For a variety of examples please cf. the EOR Web site

C.2 Recommended rules

The »encyclopedia mode« demands the following steps:
  – extraction of the strings to be ordered from the relevant data. All preparatory procedures described in the main part of this European Standard may be relevant here
  – implicit transliteration into the predominant script according to the relevant standards
  – ordering of the strings thus obtained as specified in the main part of this European Standard
  – rendering of strings in their original form, but in the order thus obtained

  To minimize the problems with user-expectations to transliteration this European Standard demands that strings be transliterated according to the relevant ISO standards. For transliteration into Latin these are *ISO R 843:1968* for Greek and *ISO R 9:1968* for Cyrillic.

Annex D (informative): Bibliography

To be added

D.1 Relevant international standards

D.2 National ordering standards

D.3 References to usage in ordering