Second draft version of the
European Ordering Rules
Date:1998.07.23

Marc Wilhelm Küster

Table of Contents

Introduction

This European Standard is intended to facilitate 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.

This European Standard must cater for two mutually exclusive demands: The implementer needs clear guidelines and data which can readily be used in existing and future ordering applications. This can best be done with reference to the forthcoming ordering standard ISO/IEC FCD 14651 of which the present standard is a profile. The user with no specific IT-background, however, needs an exposition of the principles in a form more in line with existing national ordering standards or relevant practice. As tailoring tables in the syntax of ISO/IEC FCD 14651 can be difficult to read for human readers and have no likeness to more traditional treatments of this matter, an authoritative exposition of the principles behind that table is given in the informative annexes. It is written in the style of non-IT ordering standards and users not familiar with the formal syntax of the tailoring table are advised to consult those annexes first. A web site on this subject is hosted by the Icelandic Standards Organization STRÍ for further reference.[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⟩ At present STRÍ can be accessed under `http://www.stri.is` ⟨/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 1973:1995, 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. Informative Annex A presents equivalent information in a more readily accessible way. Informative Annex B deals with word-by-word ordering as a special form of ordering with multiple keys. Informative Annex C explains the use of further ordering criteria.

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

2 Notation

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

3 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 always refer to subsequent amendments of the publication in question. Undated references always refer to the lastest 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]   Information   Technology   –   European   subsets   of ISO/IEC 10646–1:1993[4]

Unicode 2.0   The Unicode Consortium: *The Unicode Standard Version 2.0* Addison-Wesley: Reading/MA, 1996[5]

4 Definitions

For the purpose of this European standard the definitions of ISO/IEC FCD 14651:1997 apply

5 Conformance

For the purpose of this standard the conformance clause of ISO/IEC FCD 14651:1997 applies

6 Tailorability

The European Ordering Rules defined in this standard can taken as a template for tailored it to the needs of any European country in the manner specified by ISO/IEC FCD 14651

7 Default table

```
%% EOR EORTable
%% A default template for Europe: The European Ordering Rules
section EORCollatingSymbols
%% The names of the collating symbols are adapted to those chosen in Annex A
%% The syntax used is that specified in document ISO/IEC/JTC1/SC22/WG20 N578R

%% Collating symbols for the primary ordering key
%%% Digits
⟨0⟩
⟨1⟩
⟨2⟩
⟨3⟩
⟨4⟩
⟨5⟩
⟨6⟩
⟨7⟩
```

---

[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⟩

[5] ⟨NOTE⟩ At present it is not possible to reference the Unicode standard normatively. Suggestions?⟨/NOTE⟩

⟨8⟩
⟨9⟩

%%% Latin
⟨a⟩
⟨b⟩
⟨c⟩
⟨d⟩
⟨e⟩
⟨f⟩
⟨g⟩
⟨h⟩
⟨i⟩
⟨j⟩
⟨k⟩
⟨l⟩
⟨m⟩
⟨n⟩
⟨o⟩
⟨p⟩
⟨q⟩
⟨r⟩
⟨s⟩
⟨t⟩
⟨u⟩
⟨v⟩
⟨w⟩
⟨x⟩
⟨y⟩
⟨z⟩
⟨thorn⟩

%%% Greek
⟨alpha⟩
⟨beta⟩
⟨gamma⟩
⟨delta⟩
⟨epsilon⟩
⟨digamma⟩
⟨stigma⟩
⟨zeta⟩
⟨eta⟩
⟨theta⟩
⟨iota⟩
⟨kappa⟩
⟨lamda⟩
⟨mu⟩
⟨nu⟩
⟨xi⟩
⟨omicron⟩
⟨pi⟩
⟨koppa⟩
⟨rho⟩
⟨sigma⟩
⟨finalsigma⟩ %% Treated as a letter of its own at the request of ELOT
⟨tau⟩
⟨upsilon⟩
⟨phi⟩

⟨chi⟩
⟨psi⟩
⟨omega⟩

%%% Cyrillic
⟨acyril⟩
⟨be⟩
⟨ve⟩
⟨ge⟩
⟨gje⟩
⟨geupturn⟩
⟨de⟩
⟨dje⟩
⟨ie⟩
⟨io⟩
⟨ecyril⟩
⟨zhe⟩
⟨ze⟩
⟨dze⟩
⟨ii⟩
⟨icyril⟩
⟨yi⟩
⟨iibreve⟩
⟨je⟩
⟨ka⟩
⟨kje⟩
⟨el⟩
⟨lje⟩
⟨em⟩
⟨en⟩
⟨nje⟩
⟨ocyril⟩
⟨pecyril⟩
⟨er⟩
⟨es⟩
⟨te⟩
⟨tshe⟩
⟨ucyril⟩
⟨ucyrilbreve⟩
⟨ef⟩
⟨kha⟩
⟨tse⟩
⟨che⟩
⟨dzhe⟩
⟨sha⟩
⟨shcha⟩
⟨hard⟩
⟨yeri⟩
⟨soft⟩
⟨ecyrilrev⟩
⟨iu⟩
⟨ia⟩

%% Collating symbols for the secondary ordering key
⟨BLANK⟩
⟨PSILI⟩
%% This sequence follows from the treatment of ⟨OXIA⟩ and ⟨VARIA⟩
⟨PSILI+OXIA⟩

⟨PSILI+OXIA+PROSGEGRAMMENI⟩
⟨PSILI+OXIA+YPOGEGRAMMENI⟩
⟨PSILI+VARIA⟩
⟨PSILI+VARIA+PROSGEGRAMMENI⟩
⟨PSILI+VARIA+YPOGEGRAMMENI⟩
⟨PSILI+PERISPOMENI+PROSGEGRAMMENI⟩
⟨PSILI+PERISPOMENI+YPOGEGRAMMENI⟩
⟨PSILI+PERISPOMENI⟩
⟨PSILI+PROSGEGRAMMENI⟩
⟨PSILI+YPOGEGRAMMENI⟩
⟨DASIA⟩
%% This sequence follows from the treatment of ⟨OXIA⟩ and ⟨VARIA⟩
⟨DASIA+OXIA⟩
⟨DASIA+OXIA+PROSGEGRAMMENI⟩
⟨DASIA+OXIA+YPOGEGRAMMENI⟩
⟨DASIA+VARIA⟩
⟨DASIA+VARIA+PROSGEGRAMMENI⟩
⟨DASIA+VARIA+YPOGEGRAMMENI⟩
⟨DASIA+PERISPOMENI⟩
⟨DASIA+PERISPOMENI+PROSGEGRAMMENI⟩
⟨DASIA+PERISPOMENI+YPOGEGRAMMENI⟩
⟨DASIA+PROSGEGRAMMENI⟩
⟨DASIA+YPOGEGRAMMENI⟩
%% According the the Unicode rules for canonical equivalence ⟨OXIA⟩
%% is equivalent to the acute accent U0301
%% Adherence to the Unicode rules for canonical equivalence breaks thus
%% standard Greek ordering practice
⟨ACUTE⟩
⟨OXIA+YPOGEGRAMMENI⟩
%% According the the Unicode rules for canonical equivalence ⟨VARIA⟩
%% is equivalent to the grave accent U300
%% Adherence to the Unicode rules for canonical equivalence breaks thus
%% standard Greek ordering practice
⟨GRAVE⟩
⟨VARIA+YPOGEGRAMMENI⟩
⟨PERISPOMENI⟩
⟨PERISPOMENI+YPOGEGRAMMENI⟩
⟨TONOS⟩
⟨PROSGEGRAMMENI⟩
⟨YPOGEGRAMMENI⟩
⟨BREVE⟩
⟨CIRCUMFLEX⟩
⟨CARON⟩
⟨RING_ABOVE⟩
⟨RING_ABOVE+ACUTE⟩
%% ⟨DIAERESIS⟩ is unified with the Greek DIALYTICA
%% This does not influence the ordering results and allows full
%% decomposition according to the Unicode rules for canonical equivalence
⟨DIAERESIS⟩
⟨DIAERESIS+VARIA⟩
⟨DIAERESIS+OXIA⟩
⟨DIAERESIS+PERISPOMENI⟩
⟨DIAERESIS+TONOS⟩
⟨DIAERESIS+MACRON⟩
⟨DOUBLE_ACUTE⟩
⟨TILDE⟩
⟨DOT_ABOVE⟩
⟨CEDILLA⟩

⟨OGONEK⟩
%% ⟨MACRON⟩ is unified with the Greek VRACHY
%% This does not influence the ordering results and allows full
%% decomposition according to the Unicode rules for canonical equivalence
⟨MACRON⟩
⟨STROKE⟩
⟨STROKE+ACUTE⟩
⟨MODIFIED⟩
⟨MODIFIED+CARON⟩
⟨MODIFIED+APOSTROPHE⟩

%% Collating symbol for the tertiary ordering key
⟨NO_CASE⟩
⟨SMALL⟩
⟨CAPITAL⟩

section EORDigits
⟨U0030⟩ ⟨0⟩;⟨BLANK⟩;⟨NO_CASE⟩; IGNORE % DIGIT ZERO
⟨U0031⟩ ⟨1⟩;⟨BLANK⟩;⟨NO_CASE⟩; IGNORE % DIGIT ONE
⟨U0032⟩ ⟨2⟩;⟨BLANK⟩;⟨NO_CASE⟩; IGNORE % DIGIT TWO
⟨U0033⟩ ⟨3⟩;⟨BLANK⟩;⟨NO_CASE⟩; IGNORE % DIGIT THREE
⟨U0034⟩ ⟨4⟩;⟨BLANK⟩;⟨NO_CASE⟩; IGNORE % DIGIT FOUR
⟨U0035⟩ ⟨5⟩;⟨BLANK⟩;⟨NO_CASE⟩; IGNORE % DIGIT FIVE
⟨U0036⟩ ⟨6⟩;⟨BLANK⟩;⟨NO_CASE⟩; IGNORE % DIGIT SIX
⟨U0037⟩ ⟨7⟩;⟨BLANK⟩;⟨NO_CASE⟩; IGNORE % DIGIT SEVEN
⟨U0038⟩ ⟨8⟩;⟨BLANK⟩;⟨NO_CASE⟩; IGNORE % DIGIT EIGHT
⟨U0039⟩ ⟨9⟩;⟨BLANK⟩;⟨NO_CASE⟩; IGNORE % DIGIT NINE

section EORLatin
⟨U0061⟩ ⟨a⟩;⟨BLANK⟩;⟨SMALL⟩;IGNORE % LATIN SMALL LETTER A
⟨U0041⟩ ⟨a⟩;⟨BLANK⟩;⟨CAPITAL⟩;IGNORE % LATIN CAPITAL LETTER A
⟨U00E1⟩ ⟨a⟩;⟨ACUTE⟩;⟨SMALL⟩;IGNORE % LATIN SMALL LETTER A WITH ACUTE
⟨U00C1⟩ ⟨a⟩;⟨ACUTE⟩;⟨CAPITAL⟩;IGNORE % LATIN CAPITAL LETTER A WITH ACUTE
⟨U00E0⟩ ⟨a⟩;⟨GRAVE⟩;⟨SMALL⟩;IGNORE % LATIN SMALL LETTER A WITH GRAVE
⟨U00C0⟩ ⟨a⟩;⟨GRAVE⟩;⟨CAPITAL⟩;IGNORE % LATIN CAPITAL LETTER A WITH GRAVE
⟨U0103⟩ ⟨a⟩;⟨BREVE⟩;⟨SMALL⟩;IGNORE % LATIN SMALL LETTER A WITH BREVE
⟨U0102⟩ ⟨a⟩;⟨BREVE⟩;⟨CAPITAL⟩;IGNORE % LATIN CAPITAL LETTER A WITH BREVE
⟨U00E2⟩ ⟨a⟩;⟨CIRCUMFLEX⟩;⟨SMALL⟩;IGNORE % LATIN SMALL LETTER A WITH CIRCUMFLEX
⟨U00C2⟩ ⟨a⟩;⟨CIRCUMFLEX⟩;⟨CAPITAL⟩;IGNORE % LATIN CAPITAL LETTER A WITH CIRCUMFLEX
⟨U00E5⟩ ⟨a⟩;⟨RING_ABOVE⟩;⟨SMALL⟩;IGNORE % LATIN SMALL LETTER A WITH RING ABOVE
⟨U00C5⟩ ⟨a⟩;⟨RING_ABOVE⟩;⟨CAPITAL⟩;IGNORE % LATIN CAPITAL LETTER A WITH RING ABOVE
⟨U01FB⟩ ⟨a⟩;⟨RING_ABOVE+ACUTE⟩;⟨SMALL⟩;IGNORE % LATIN SMALL LETTER A WITH RING ABOVE AND ACUTE
⟨U01FA⟩ ⟨a⟩;⟨RING_ABOVE+ACUTE⟩;⟨CAPITAL⟩;IGNORE % LATIN CAPITAL LETTER A WITH RING ABOVE AND ACUTE
⟨U00E4⟩ ⟨a⟩;⟨DIAERESIS⟩;⟨SMALL⟩;IGNORE % LATIN SMALL LETTER A WITH DIAERESIS
⟨U00C4⟩ ⟨a⟩;⟨DIAERESIS⟩;⟨CAPITAL⟩;IGNORE % LATIN CAPITAL LETTER A WITH DIAERESIS
⟨U01DF⟩ ⟨a⟩;⟨DIAERESIS+MACRON⟩;⟨SMALL⟩;IGNORE % LATIN SMALL LETTER A WITH DIAERESIS AND MACRON
⟨U01DE⟩ ⟨a⟩;⟨DIAERESIS+MACRON⟩;⟨CAPITAL⟩;IGNORE % LATIN CAPITAL LETTER A WITH DIAERESIS AND MACRON
⟨U00E3⟩ ⟨a⟩;⟨TILDE⟩;⟨SMALL⟩;IGNORE % LATIN SMALL LETTER A WITH TILDE
⟨U00C3⟩ ⟨a⟩;⟨TILDE⟩;⟨CAPITAL⟩;IGNORE % LATIN CAPITAL LETTER A WITH TILDE
⟨U0105⟩ ⟨a⟩;⟨OGONEK⟩;⟨SMALL⟩;IGNORE % LATIN SMALL LETTER A WITH OGONEK
⟨U0104⟩ ⟨a⟩;⟨OGONEK⟩;⟨CAPITAL⟩;IGNORE % LATIN CAPITAL LETTER A WITH OGONEK
⟨U0101⟩ ⟨a⟩;⟨MACRON⟩;⟨SMALL⟩;IGNORE % LATIN SMALL LETTER A WITH MACRON
⟨U0100⟩ ⟨a⟩;⟨MACRON⟩;⟨CAPITAL⟩;IGNORE % LATIN CAPITAL LETTER A WITH MACRON
⟨U00E6⟩ "⟨a⟩⟨e⟩";"⟨BLANK⟩⟨BLANK⟩";"⟨SMALL⟩⟨SMALL⟩";IGNORE % LATIN SMALL LETTER AE

7

⟨U00C6⟩ ″⟨a⟩⟨e⟩″;″⟨BLANK⟩⟨BLANK⟩″;″⟨CAPITAL⟩⟨CAPITAL⟩″;IGNORE % LATIN CAPITAL LETTER AE
⟨U01FD⟩ ″⟨a⟩⟨e⟩″;″⟨ACUTE⟩⟨BLANK⟩″;″⟨SMALL⟩⟨SMALL⟩″;IGNORE % LATIN SMALL LETTER AE WITH ACUTE
⟨U01FC⟩ ″⟨a⟩⟨e⟩″;″⟨ACUTE⟩⟨BLANK⟩″;″⟨CAPITAL⟩⟨CAPITAL⟩″;IGNORE % LATIN CAPITAL LETTER AE WITH ACUTE
⟨U0062⟩ ⟨b⟩;⟨BLANK⟩;⟨SMALL⟩;IGNORE % LATIN SMALL LETTER B
⟨U0042⟩ ⟨b⟩;⟨BLANK⟩;⟨CAPITAL⟩;IGNORE % LATIN CAPITAL LETTER B
⟨U1E03⟩ ⟨b⟩;⟨DOT_ABOVE⟩;⟨SMALL⟩;IGNORE % LATIN SMALL LETTER B WITH DOT ABOVE
⟨U1E02⟩ ⟨b⟩;⟨DOT_ABOVE⟩;⟨CAPITAL⟩;IGNORE % LATIN CAPITAL LETTER B WITH DOT ABOVE
⟨U0063⟩ ⟨c⟩;⟨BLANK⟩;⟨SMALL⟩;IGNORE % LATIN SMALL LETTER C
⟨U0043⟩ ⟨c⟩;⟨BLANK⟩;⟨CAPITAL⟩;IGNORE % LATIN CAPITAL LETTER C
⟨U0107⟩ ⟨c⟩;⟨ACUTE⟩;⟨SMALL⟩;IGNORE % LATIN SMALL LETTER C WITH ACUTE
⟨U0106⟩ ⟨c⟩;⟨ACUTE⟩;⟨CAPITAL⟩;IGNORE % LATIN CAPITAL LETTER C WITH ACUTE
⟨U0109⟩ ⟨c⟩;⟨CIRCUMFLEX⟩;⟨SMALL⟩;IGNORE % LATIN SMALL LETTER C WITH CIRCUMFLEX
⟨U0108⟩ ⟨c⟩;⟨CIRCUMFLEX⟩;⟨CAPITAL⟩;IGNORE % LATIN CAPITAL LETTER C WITH CIRCUMFLEX
⟨U010D⟩ ⟨c⟩;⟨CARON⟩;⟨SMALL⟩;IGNORE % LATIN SMALL LETTER C WITH CARON
⟨U010C⟩ ⟨c⟩;⟨CARON⟩;⟨CAPITAL⟩;IGNORE % LATIN CAPITAL LETTER C WITH CARON
⟨U010B⟩ ⟨c⟩;⟨DOT_ABOVE⟩;⟨SMALL⟩;IGNORE % LATIN SMALL LETTER C WITH DOT ABOVE
⟨U010A⟩ ⟨c⟩;⟨DOT_ABOVE⟩;⟨CAPITAL⟩;IGNORE % LATIN CAPITAL LETTER C WITH DOT ABOVE
⟨U00E7⟩ ⟨c⟩;⟨CEDILLA⟩;⟨SMALL⟩;IGNORE % LATIN SMALL LETTER C WITH CEDILLA
⟨U00C7⟩ ⟨c⟩;⟨CEDILLA⟩;⟨CAPITAL⟩;IGNORE % LATIN CAPITAL LETTER C WITH CEDILLA
⟨U0064⟩ ⟨d⟩;⟨BLANK⟩;⟨SMALL⟩;IGNORE % LATIN SMALL LETTER D
⟨U0044⟩ ⟨d⟩;⟨BLANK⟩;⟨CAPITAL⟩;IGNORE % LATIN CAPITAL LETTER D
⟨U010F⟩ ⟨d⟩;⟨CARON⟩;⟨SMALL⟩;IGNORE % LATIN SMALL LETTER D WITH CARON
⟨U010E⟩ ⟨d⟩;⟨CARON⟩;⟨CAPITAL⟩;IGNORE % LATIN CAPITAL LETTER D WITH CARON
⟨U1E0B⟩ ⟨d⟩;⟨DOT_ABOVE⟩;⟨SMALL⟩;IGNORE % LATIN SMALL LETTER D WITH DOT ABOVE
⟨U1E0A⟩ ⟨d⟩;⟨DOT_ABOVE⟩;⟨CAPITAL⟩;IGNORE % LATIN CAPITAL LETTER D WITH DOT ABOVE
⟨U1E11⟩ ⟨d⟩;⟨CEDILLA⟩;⟨SMALL⟩;IGNORE % LATIN SMALL LETTER D WITH CEDILLA
⟨U1E10⟩ ⟨d⟩;⟨CEDILLA⟩;⟨CAPITAL⟩;IGNORE % LATIN CAPITAL LETTER D WITH CEDILLA
⟨U0111⟩ ⟨d⟩;⟨STROKE⟩;⟨SMALL⟩;IGNORE % LATIN SMALL LETTER D WITH STROKE
⟨U0110⟩ ⟨d⟩;⟨STROKE⟩;⟨CAPITAL⟩;IGNORE % LATIN CAPITAL LETTER D WITH STROKE
⟨U00F0⟩ ⟨d⟩;⟨MODIFIED⟩;⟨SMALL⟩;IGNORE % LATIN SMALL LETTER ETH
⟨U00D0⟩ ⟨d⟩;⟨MODIFIED⟩;⟨CAPITAL⟩;IGNORE % LATIN CAPITAL LETTER ETH
⟨U01F3⟩ ″⟨d⟩⟨z⟩″;″⟨BLANK⟩⟨BLANK⟩″;″⟨SMALL⟩⟨SMALL⟩″;IGNORE % LATIN SMALL LETTER DZ
⟨U01F1⟩ ″⟨d⟩⟨z⟩″;″⟨BLANK⟩⟨BLANK⟩″;″⟨CAPITAL⟩⟨CAPITAL⟩″;IGNORE % LATIN CAPITAL LETTER DZ
⟨U01C6⟩ ″⟨d⟩⟨z⟩″;″⟨BLANK⟩⟨CARON⟩″;″⟨SMALL⟩⟨SMALL⟩″;IGNORE % LATIN SMALL LETTER DZ WITH CARON
⟨U01C4⟩ ″⟨d⟩⟨z⟩″;″⟨BLANK⟩⟨CARON⟩″;″⟨CAPITAL⟩⟨CAPITAL⟩″;IGNORE % LATIN CAPITAL LETTER DZ WITH CARON
⟨U0065⟩ ⟨e⟩;⟨BLANK⟩;⟨SMALL⟩;IGNORE % LATIN SMALL LETTER E
⟨U0045⟩ ⟨e⟩;⟨BLANK⟩;⟨CAPITAL⟩;IGNORE % LATIN CAPITAL LETTER E
⟨U00E9⟩ ⟨e⟩;⟨ACUTE⟩;⟨SMALL⟩;IGNORE % LATIN SMALL LETTER E WITH ACUTE
⟨U00C9⟩ ⟨e⟩;⟨ACUTE⟩;⟨CAPITAL⟩;IGNORE % LATIN CAPITAL LETTER E WITH ACUTE
⟨U00E8⟩ ⟨e⟩;⟨GRAVE⟩;⟨SMALL⟩;IGNORE % LATIN SMALL LETTER E WITH GRAVE
⟨U00C8⟩ ⟨e⟩;⟨GRAVE⟩;⟨CAPITAL⟩;IGNORE % LATIN CAPITAL LETTER E WITH GRAVE
⟨U0115⟩ ⟨e⟩;⟨BREVE⟩;⟨SMALL⟩;IGNORE % LATIN SMALL LETTER E WITH BREVE
⟨U0114⟩ ⟨e⟩;⟨BREVE⟩;⟨CAPITAL⟩;IGNORE % LATIN CAPITAL LETTER E WITH BREVE
⟨U00EA⟩ ⟨e⟩;⟨CIRCUMFLEX⟩;⟨SMALL⟩;IGNORE % LATIN SMALL LETTER E WITH CIRCUMFLEX
⟨U00CA⟩ ⟨e⟩;⟨CIRCUMFLEX⟩;⟨CAPITAL⟩;IGNORE % LATIN CAPITAL LETTER E WITH CIRCUMFLEX
⟨U011B⟩ ⟨e⟩;⟨CARON⟩;⟨SMALL⟩;IGNORE % LATIN SMALL LETTER E WITH CARON
⟨U011A⟩ ⟨e⟩;⟨CARON⟩;⟨CAPITAL⟩;IGNORE % LATIN CAPITAL LETTER E WITH CARON
⟨U00EB⟩ ⟨e⟩;⟨DIAERESIS⟩;⟨SMALL⟩;IGNORE % LATIN SMALL LETTER E WITH DIAERESIS
⟨U00CB⟩ ⟨e⟩;⟨DIAERESIS⟩;⟨CAPITAL⟩;IGNORE % LATIN CAPITAL LETTER E WITH DIAERESIS
⟨U0117⟩ ⟨e⟩;⟨DOT_ABOVE⟩;⟨SMALL⟩;IGNORE % LATIN SMALL LETTER E WITH DOT ABOVE
⟨U0116⟩ ⟨e⟩;⟨DOT_ABOVE⟩;⟨CAPITAL⟩;IGNORE % LATIN CAPITAL LETTER E WITH DOT ABOVE
⟨U0119⟩ ⟨e⟩;⟨OGONEK⟩;⟨SMALL⟩;IGNORE % LATIN SMALL LETTER E WITH OGONEK

8

⟨U0118⟩ ⟨e⟩;⟨OGONEK⟩;⟨CAPITAL⟩;IGNORE % LATIN CAPITAL LETTER E WITH OGONEK
⟨U0113⟩ ⟨e⟩;⟨MACRON⟩;⟨SMALL⟩;IGNORE % LATIN SMALL LETTER E WITH MACRON
⟨U0112⟩ ⟨e⟩;⟨MACRON⟩;⟨CAPITAL⟩;IGNORE % LATIN CAPITAL LETTER E WITH MACRON
⟨U0066⟩ ⟨f⟩;⟨BLANK⟩;⟨SMALL⟩;IGNORE % LATIN SMALL LETTER F
⟨U0046⟩ ⟨f⟩;⟨BLANK⟩;⟨CAPITAL⟩;IGNORE % LATIN CAPITAL LETTER F
⟨U1E1F⟩ ⟨f⟩;⟨DOT_ABOVE⟩;⟨SMALL⟩;IGNORE % LATIN SMALL LETTER F WITH DOT ABOVE
⟨U1E1E⟩ ⟨f⟩;⟨DOT_ABOVE⟩;⟨CAPITAL⟩;IGNORE % LATIN CAPITAL LETTER F WITH DOT ABOVE
⟨U0192⟩ ⟨f⟩;⟨MODIFIED⟩;⟨SMALL⟩;IGNORE % LATIN SMALL LETTER F WITH HOOK
⟨UFB01⟩ "⟨f⟩⟨i⟩";"⟨BLANK⟩⟨BLANK⟩";"⟨SMALL⟩⟨SMALL⟩";IGNORE % LATIN SMALL LIGATURE FI
⟨UFB02⟩ "⟨f⟩⟨l⟩";"⟨BLANK⟩⟨BLANK⟩";"⟨SMALL⟩⟨SMALL⟩";IGNORE % LATIN SMALL LIGATURE FL
⟨U0067⟩ ⟨g⟩;⟨BLANK⟩;⟨SMALL⟩;IGNORE % LATIN SMALL LETTER G
⟨U0047⟩ ⟨g⟩;⟨BLANK⟩;⟨CAPITAL⟩;IGNORE % LATIN CAPITAL LETTER G
⟨U01F5⟩ ⟨g⟩;⟨ACUTE⟩;⟨SMALL⟩;IGNORE % LATIN SMALL LETTER G WITH ACUTE
⟨U01F4⟩ ⟨g⟩;⟨ACUTE⟩;⟨CAPITAL⟩;IGNORE % LATIN CAPITAL LETTER G WITH ACUTE
⟨U011F⟩ ⟨g⟩;⟨BREVE⟩;⟨SMALL⟩;IGNORE % LATIN SMALL LETTER G WITH BREVE
⟨U011E⟩ ⟨g⟩;⟨BREVE⟩;⟨CAPITAL⟩;IGNORE % LATIN CAPITAL LETTER G WITH BREVE
⟨U011D⟩ ⟨g⟩;⟨CIRCUMFLEX⟩;⟨SMALL⟩;IGNORE % LATIN SMALL LETTER G WITH CIRCUMFLEX
⟨U011C⟩ ⟨g⟩;⟨CIRCUMFLEX⟩;⟨CAPITAL⟩;IGNORE % LATIN CAPITAL LETTER G WITH CIRCUMFLEX
⟨U01E7⟩ ⟨g⟩;⟨CARON⟩;⟨SMALL⟩;IGNORE % LATIN SMALL LETTER G WITH CARON
⟨U01E6⟩ ⟨g⟩;⟨CARON⟩;⟨CAPITAL⟩;IGNORE % LATIN CAPITAL LETTER G WITH CARON
⟨U0121⟩ ⟨g⟩;⟨DOT_ABOVE⟩;⟨SMALL⟩;IGNORE % LATIN SMALL LETTER G WITH DOT ABOVE
⟨U0120⟩ ⟨g⟩;⟨DOT_ABOVE⟩;⟨CAPITAL⟩;IGNORE % LATIN CAPITAL LETTER G WITH DOT ABOVE
⟨U0123⟩ ⟨g⟩;⟨CEDILLA⟩;⟨SMALL⟩;IGNORE % LATIN SMALL LETTER G WITH CEDILLA
⟨U0122⟩ ⟨g⟩;⟨CEDILLA⟩;⟨CAPITAL⟩;IGNORE % LATIN CAPITAL LETTER G WITH CEDILLA
⟨U01E5⟩ ⟨g⟩;⟨STROKE⟩;⟨SMALL⟩;IGNORE % LATIN SMALL LETTER G WITH STROKE
⟨U01E4⟩ ⟨g⟩;⟨STROKE⟩;⟨CAPITAL⟩;IGNORE % LATIN CAPITAL LETTER G WITH STROKE
⟨U0068⟩ ⟨h⟩;⟨BLANK⟩;⟨SMALL⟩;IGNORE % LATIN SMALL LETTER H
⟨U0048⟩ ⟨h⟩;⟨BLANK⟩;⟨CAPITAL⟩;IGNORE % LATIN CAPITAL LETTER H
⟨U0125⟩ ⟨h⟩;⟨CIRCUMFLEX⟩;⟨SMALL⟩;IGNORE % LATIN SMALL LETTER H WITH CIRCUMFLEX
⟨U0124⟩ ⟨h⟩;⟨CIRCUMFLEX⟩;⟨CAPITAL⟩;IGNORE % LATIN CAPITAL LETTER H WITH CIRCUMFLEX
⟨U0127⟩ ⟨h⟩;⟨STROKE⟩;⟨SMALL⟩;IGNORE % LATIN SMALL LETTER H WITH STROKE
⟨U0126⟩ ⟨h⟩;⟨STROKE⟩;⟨CAPITAL⟩;IGNORE % LATIN CAPITAL LETTER H WITH STROKE
⟨U0069⟩ ⟨i⟩;⟨BLANK⟩;⟨SMALL⟩;IGNORE % LATIN SMALL LETTER I
⟨U0049⟩ ⟨i⟩;⟨BLANK⟩;⟨CAPITAL⟩;IGNORE % LATIN CAPITAL LETTER I
⟨U00ED⟩ ⟨i⟩;⟨ACUTE⟩;⟨SMALL⟩;IGNORE % LATIN SMALL LETTER I WITH ACUTE
⟨U00CD⟩ ⟨i⟩;⟨ACUTE⟩;⟨CAPITAL⟩;IGNORE % LATIN CAPITAL LETTER I WITH ACUTE
⟨U00EC⟩ ⟨i⟩;⟨GRAVE⟩;⟨SMALL⟩;IGNORE % LATIN SMALL LETTER I WITH GRAVE
⟨U00CC⟩ ⟨i⟩;⟨GRAVE⟩;⟨CAPITAL⟩;IGNORE % LATIN CAPITAL LETTER I WITH GRAVE
⟨U012D⟩ ⟨i⟩;⟨BREVE⟩;⟨SMALL⟩;IGNORE % LATIN SMALL LETTER I WITH BREVE
⟨U012C⟩ ⟨i⟩;⟨BREVE⟩;⟨CAPITAL⟩;IGNORE % LATIN CAPITAL LETTER I WITH BREVE
⟨U00EE⟩ ⟨i⟩;⟨CIRCUMFLEX⟩;⟨SMALL⟩;IGNORE % LATIN SMALL LETTER I WITH CIRCUMFLEX
⟨U00CE⟩ ⟨i⟩;⟨CIRCUMFLEX⟩;⟨CAPITAL⟩;IGNORE % LATIN CAPITAL LETTER I WITH CIRCUMFLEX
⟨U00EF⟩ ⟨i⟩;⟨DIAERESIS⟩;⟨SMALL⟩;IGNORE % LATIN SMALL LETTER I WITH DIAERESIS
⟨U00CF⟩ ⟨i⟩;⟨DIAERESIS⟩;⟨CAPITAL⟩;IGNORE % LATIN CAPITAL LETTER I WITH DIAERESIS
⟨U0129⟩ ⟨i⟩;⟨TILDE⟩;⟨SMALL⟩;IGNORE % LATIN SMALL LETTER I WITH TILDE
⟨U0128⟩ ⟨i⟩;⟨TILDE⟩;⟨CAPITAL⟩;IGNORE % LATIN CAPITAL LETTER I WITH TILDE
⟨U012F⟩ ⟨i⟩;⟨OGONEK⟩;⟨SMALL⟩;IGNORE % LATIN SMALL LETTER I WITH OGONEK
⟨U012E⟩ ⟨i⟩;⟨OGONEK⟩;⟨CAPITAL⟩;IGNORE % LATIN CAPITAL LETTER I WITH OGONEK
⟨U012B⟩ ⟨i⟩;⟨MACRON⟩;⟨SMALL⟩;IGNORE % LATIN SMALL LETTER I WITH MACRON
⟨U012A⟩ ⟨i⟩;⟨MACRON⟩;⟨CAPITAL⟩;IGNORE % LATIN CAPITAL LETTER I WITH MACRON
⟨U0131⟩ ⟨i⟩;⟨MODIFIED⟩;⟨SMALL⟩;IGNORE % LATIN SMALL LETTER DOTLESS I
⟨U0130⟩ ⟨i⟩;⟨MODIFIED⟩;⟨CAPITAL⟩;IGNORE % LATIN CAPITAL LETTER I WITH DOT ABOVE
⟨U0133⟩ "⟨i⟩⟨j⟩";"⟨BLANK⟩⟨BLANK⟩";"⟨SMALL⟩⟨SMALL⟩";IGNORE % LATIN SMALL LIGATURE IJ
⟨U0132⟩ "⟨i⟩⟨j⟩";"⟨BLANK⟩⟨BLANK⟩";"⟨CAPITAL⟩⟨CAPITAL⟩";IGNORE % LATIN CAPITAL LIGATURE IJ
⟨U006A⟩ ⟨j⟩;⟨BLANK⟩;⟨SMALL⟩;IGNORE % LATIN SMALL LETTER J
⟨U004A⟩ ⟨j⟩;⟨BLANK⟩;⟨CAPITAL⟩;IGNORE % LATIN CAPITAL LETTER J
⟨U0135⟩ ⟨j⟩;⟨CIRCUMFLEX⟩;⟨SMALL⟩;IGNORE % LATIN SMALL LETTER J WITH CIRCUMFLEX

⟨U0134⟩ ⟨j⟩;⟨CIRCUMFLEX⟩;⟨CAPITAL⟩;IGNORE % LATIN CAPITAL LETTER J WITH CIRCUMFLEX
⟨U006B⟩ ⟨k⟩;⟨BLANK⟩;⟨SMALL⟩;IGNORE % LATIN SMALL LETTER K
⟨U004B⟩ ⟨k⟩;⟨BLANK⟩;⟨CAPITAL⟩;IGNORE % LATIN CAPITAL LETTER K
⟨U1E31⟩ ⟨k⟩;⟨ACUTE⟩;⟨SMALL⟩;IGNORE % LATIN SMALL LETTER K WITH ACUTE
⟨U1E30⟩ ⟨k⟩;⟨ACUTE⟩;⟨CAPITAL⟩;IGNORE % LATIN CAPITAL LETTER K WITH ACUTE
⟨U01E9⟩ ⟨k⟩;⟨CARON⟩;⟨SMALL⟩;IGNORE % LATIN SMALL LETTER K WITH CARON
⟨U01E8⟩ ⟨k⟩;⟨CARON⟩;⟨CAPITAL⟩;IGNORE % LATIN CAPITAL LETTER K WITH CARON
⟨U0137⟩ ⟨k⟩;⟨CEDILLA⟩;⟨SMALL⟩;IGNORE % LATIN SMALL LETTER K WITH CEDILLA
⟨U0136⟩ ⟨k⟩;⟨CEDILLA⟩;⟨CAPITAL⟩;IGNORE % LATIN CAPITAL LETTER K WITH CEDILLA
⟨U0138⟩ ⟨k⟩;⟨MODIFIED⟩;⟨SMALL⟩;IGNORE % LATIN SMALL LETTER KRA
⟨U006C⟩ ⟨l⟩;⟨BLANK⟩;⟨SMALL⟩;IGNORE % LATIN SMALL LETTER L
⟨U004C⟩ ⟨l⟩;⟨BLANK⟩;⟨CAPITAL⟩;IGNORE % LATIN CAPITAL LETTER L
⟨U013A⟩ ⟨l⟩;⟨ACUTE⟩;⟨SMALL⟩;IGNORE % LATIN SMALL LETTER L WITH ACUTE
⟨U0139⟩ ⟨l⟩;⟨ACUTE⟩;⟨CAPITAL⟩;IGNORE % LATIN CAPITAL LETTER L WITH ACUTE
⟨U013E⟩ ⟨l⟩;⟨CARON⟩;⟨SMALL⟩;IGNORE % LATIN SMALL LETTER L WITH CARON
⟨U013D⟩ ⟨l⟩;⟨CARON⟩;⟨CAPITAL⟩;IGNORE % LATIN CAPITAL LETTER L WITH CARON
⟨U013C⟩ ⟨l⟩;⟨CEDILLA⟩;⟨SMALL⟩;IGNORE % LATIN SMALL LETTER L WITH CEDILLA
⟨U013B⟩ ⟨l⟩;⟨CEDILLA⟩;⟨CAPITAL⟩;IGNORE % LATIN CAPITAL LETTER L WITH CEDILLA
⟨U0142⟩ ⟨l⟩;⟨STROKE⟩;⟨SMALL⟩;IGNORE % LATIN SMALL LETTER L WITH STROKE
⟨U0141⟩ ⟨l⟩;⟨STROKE⟩;⟨CAPITAL⟩;IGNORE % LATIN CAPITAL LETTER L WITH STROKE
⟨U0140⟩ ⟨l⟩;⟨MODIFIED⟩;⟨SMALL⟩;IGNORE % LATIN SMALL LETTER L WITH MIDDLE DOT
⟨U013F⟩ ⟨l⟩;⟨MODIFIED⟩;⟨CAPITAL⟩;IGNORE % LATIN CAPITAL LETTER L WITH MIDDLE DOT
⟨U01C9⟩ ”⟨l⟩⟨j⟩”;”⟨BLANK⟩⟨BLANK⟩”;”⟨SMALL⟩⟨SMALL⟩”;IGNORE % LATIN SMALL LETTER LJ
⟨U01C7⟩ ”⟨l⟩⟨j⟩”;”⟨BLANK⟩⟨BLANK⟩”;”⟨CAPITAL⟩⟨CAPITAL⟩”;IGNORE % LATIN CAPITAL LETTER LJ
⟨U006D⟩ ⟨m⟩;⟨BLANK⟩;⟨SMALL⟩;IGNORE % LATIN SMALL LETTER M
⟨U004D⟩ ⟨m⟩;⟨BLANK⟩;⟨CAPITAL⟩;IGNORE % LATIN CAPITAL LETTER M
⟨U1E41⟩ ⟨m⟩;⟨DOT_ABOVE⟩;⟨SMALL⟩;IGNORE % LATIN SMALL LETTER M WITH DOT ABOVE
⟨U1E40⟩ ⟨m⟩;⟨DOT_ABOVE⟩;⟨CAPITAL⟩;IGNORE % LATIN CAPITAL LETTER M WITH DOT ABOVE
⟨U006E⟩ ⟨n⟩;⟨BLANK⟩;⟨SMALL⟩;IGNORE % LATIN SMALL LETTER N
⟨U004E⟩ ⟨n⟩;⟨BLANK⟩;⟨CAPITAL⟩;IGNORE % LATIN CAPITAL LETTER N
⟨U0144⟩ ⟨n⟩;⟨ACUTE⟩;⟨SMALL⟩;IGNORE % LATIN SMALL LETTER N WITH ACUTE
⟨U0143⟩ ⟨n⟩;⟨ACUTE⟩;⟨CAPITAL⟩;IGNORE % LATIN CAPITAL LETTER N WITH ACUTE
⟨U0148⟩ ⟨n⟩;⟨CARON⟩;⟨SMALL⟩;IGNORE % LATIN SMALL LETTER N WITH CARON
⟨U0147⟩ ⟨n⟩;⟨CARON⟩;⟨CAPITAL⟩;IGNORE % LATIN CAPITAL LETTER N WITH CARON
⟨U00F1⟩ ⟨n⟩;⟨TILDE⟩;⟨SMALL⟩;IGNORE % LATIN SMALL LETTER N WITH TILDE
⟨U00D1⟩ ⟨n⟩;⟨TILDE⟩;⟨CAPITAL⟩;IGNORE % LATIN CAPITAL LETTER N WITH TILDE
⟨U0146⟩ ⟨n⟩;⟨CEDILLA⟩;⟨SMALL⟩;IGNORE % LATIN SMALL LETTER N WITH CEDILLA
⟨U0145⟩ ⟨n⟩;⟨CEDILLA⟩;⟨CAPITAL⟩;IGNORE % LATIN CAPITAL LETTER N WITH CEDILLA
⟨U207F⟩ ⟨n⟩;⟨MODIFIED⟩;⟨SMALL⟩;IGNORE % SUPERSCRIPT LATIN SMALL LETTER N
⟨U0149⟩ ⟨n⟩;⟨MODIFIED+APOSTROPHE⟩;⟨SMALL⟩;IGNORE % LATIN SMALL LETTER N PRECEDED BY APOSTROPHE
⟨U014B⟩ ”⟨n⟩⟨g⟩”;”⟨MODIFIED⟩⟨MODIFIED⟩”;”⟨SMALL⟩⟨SMALL⟩”;IGNORE % LATIN SMALL LETTER ENG
⟨U014A⟩ ”⟨n⟩⟨g⟩”;”⟨MODIFIED⟩⟨MODIFIED⟩”;”⟨CAPITAL⟩⟨CAPITAL⟩”;IGNORE % LATIN CAPITAL LETTER ENG
⟨U01CC⟩ ”⟨n⟩⟨j⟩”;”⟨BLANK⟩⟨BLANK⟩”;”⟨SMALL⟩⟨SMALL⟩”;IGNORE % LATIN SMALL LETTER NJ
⟨U01CA⟩ ”⟨n⟩⟨j⟩”;”⟨BLANK⟩⟨BLANK⟩”;”⟨CAPITAL⟩⟨CAPITAL⟩”;IGNORE % LATIN CAPITAL LETTER NJ
⟨U006F⟩ ⟨o⟩;⟨BLANK⟩;⟨SMALL⟩;IGNORE % LATIN SMALL LETTER O
⟨U004F⟩ ⟨o⟩;⟨BLANK⟩;⟨CAPITAL⟩;IGNORE % LATIN CAPITAL LETTER O
⟨U00F3⟩ ⟨o⟩;⟨ACUTE⟩;⟨SMALL⟩;IGNORE % LATIN SMALL LETTER O WITH ACUTE
⟨U00D3⟩ ⟨o⟩;⟨ACUTE⟩;⟨CAPITAL⟩;IGNORE % LATIN CAPITAL LETTER O WITH ACUTE
⟨U00F2⟩ ⟨o⟩;⟨GRAVE⟩;⟨SMALL⟩;IGNORE % LATIN SMALL LETTER O WITH GRAVE
⟨U00D2⟩ ⟨o⟩;⟨GRAVE⟩;⟨CAPITAL⟩;IGNORE % LATIN CAPITAL LETTER O WITH GRAVE
⟨U014F⟩ ⟨o⟩;⟨BREVE⟩;⟨SMALL⟩;IGNORE % LATIN SMALL LETTER O WITH BREVE
⟨U014E⟩ ⟨o⟩;⟨BREVE⟩;⟨CAPITAL⟩;IGNORE % LATIN CAPITAL LETTER O WITH BREVE
⟨U00F4⟩ ⟨o⟩;⟨CIRCUMFLEX⟩;⟨SMALL⟩;IGNORE % LATIN SMALL LETTER O WITH CIRCUMFLEX

⟨U00D4⟩ ⟨o⟩;⟨CIRCUMFLEX⟩;⟨CAPITAL⟩;IGNORE % LATIN CAPITAL LETTER O WITH CIRCUMFLEX
⟨U00F6⟩ ⟨o⟩;⟨DIAERESIS⟩;⟨SMALL⟩;IGNORE % LATIN SMALL LETTER O WITH DIAERESIS
⟨U00D6⟩ ⟨o⟩;⟨DIAERESIS⟩;⟨CAPITAL⟩;IGNORE % LATIN CAPITAL LETTER O WITH DIAERESIS
⟨U0151⟩ ⟨o⟩;⟨DOUBLE_ACUTE⟩;⟨SMALL⟩;IGNORE % LATIN SMALL LETTER O WITH DOUBLE ACUTE
⟨U0150⟩ ⟨o⟩;⟨DOUBLE_ACUTE⟩;⟨CAPITAL⟩;IGNORE % LATIN CAPITAL LETTER O WITH DOUBLE ACUTE
⟨U00F5⟩ ⟨o⟩;⟨TILDE⟩;⟨SMALL⟩;IGNORE % LATIN SMALL LETTER O WITH TILDE
⟨U00D5⟩ ⟨o⟩;⟨TILDE⟩;⟨CAPITAL⟩;IGNORE % LATIN CAPITAL LETTER O WITH TILDE
⟨U014D⟩ ⟨o⟩;⟨MACRON⟩;⟨SMALL⟩;IGNORE % LATIN SMALL LETTER O WITH MACRON
⟨U014C⟩ ⟨o⟩;⟨MACRON⟩;⟨CAPITAL⟩;IGNORE % LATIN CAPITAL LETTER O WITH MACRON
⟨U00F8⟩ ⟨o⟩;⟨STROKE⟩;⟨SMALL⟩;IGNORE % LATIN SMALL LETTER O WITH STROKE
⟨U00D8⟩ ⟨o⟩;⟨STROKE⟩;⟨CAPITAL⟩;IGNORE % LATIN CAPITAL LETTER O WITH STROKE
⟨U01FF⟩ ⟨o⟩;⟨STROKE+ACUTE⟩;⟨SMALL⟩;IGNORE % LATIN SMALL LETTER O WITH STROKE AND ACUTE
⟨U01FE⟩ ⟨o⟩;⟨STROKE+ACUTE⟩;⟨CAPITAL⟩;IGNORE % LATIN CAPITAL LETTER O WITH STROKE AND ACUTE
⟨U0153⟩ ″⟨o⟩⟨e⟩″;″⟨BLANK⟩⟨BLANK⟩″;″⟨SMALL⟩⟨SMALL⟩″;IGNORE % LATIN SMALL LIGATURE OE
⟨U0152⟩ ″⟨o⟩⟨e⟩″;″⟨BLANK⟩⟨BLANK⟩″;″⟨CAPITAL⟩⟨CAPITAL⟩″;IGNORE % LATIN CAPITAL LIGATURE OE
⟨U0070⟩ ⟨p⟩;⟨BLANK⟩;⟨SMALL⟩;IGNORE % LATIN SMALL LETTER P
⟨U0050⟩ ⟨p⟩;⟨BLANK⟩;⟨CAPITAL⟩;IGNORE % LATIN CAPITAL LETTER P
⟨U1E57⟩ ⟨p⟩;⟨DOT_ABOVE⟩;⟨SMALL⟩;IGNORE % LATIN SMALL LETTER P WITH DOT ABOVE
⟨U1E56⟩ ⟨p⟩;⟨DOT_ABOVE⟩;⟨CAPITAL⟩;IGNORE % LATIN CAPITAL LETTER P WITH DOT ABOVE
⟨U0071⟩ ⟨q⟩;⟨BLANK⟩;⟨SMALL⟩;IGNORE % LATIN SMALL LETTER Q
⟨U0051⟩ ⟨q⟩;⟨BLANK⟩;⟨CAPITAL⟩;IGNORE % LATIN CAPITAL LETTER Q
⟨U0072⟩ ⟨r⟩;⟨BLANK⟩;⟨SMALL⟩;IGNORE % LATIN SMALL LETTER R
⟨U0052⟩ ⟨r⟩;⟨BLANK⟩;⟨CAPITAL⟩;IGNORE % LATIN CAPITAL LETTER R
⟨U0155⟩ ⟨r⟩;⟨ACUTE⟩;⟨SMALL⟩;IGNORE % LATIN SMALL LETTER R WITH ACUTE
⟨U0154⟩ ⟨r⟩;⟨ACUTE⟩;⟨CAPITAL⟩;IGNORE % LATIN CAPITAL LETTER R WITH ACUTE
⟨U0159⟩ ⟨r⟩;⟨CARON⟩;⟨SMALL⟩;IGNORE % LATIN SMALL LETTER R WITH CARON
⟨U0158⟩ ⟨r⟩;⟨CARON⟩;⟨CAPITAL⟩;IGNORE % LATIN CAPITAL LETTER R WITH CARON
⟨U0157⟩ ⟨r⟩;⟨CEDILLA⟩;⟨SMALL⟩;IGNORE % LATIN SMALL LETTER R WITH CEDILLA
⟨U0156⟩ ⟨r⟩;⟨CEDILLA⟩;⟨CAPITAL⟩;IGNORE % LATIN CAPITAL LETTER R WITH CEDILLA
⟨U027C⟩ ⟨r⟩;⟨MODIFIED⟩;⟨SMALL⟩;IGNORE % LATIN SMALL LETTER R WITH LONG LEG
⟨U0073⟩ ⟨s⟩;⟨BLANK⟩;⟨SMALL⟩;IGNORE % LATIN SMALL LETTER S
⟨U0053⟩ ⟨s⟩;⟨BLANK⟩;⟨CAPITAL⟩;IGNORE % LATIN CAPITAL LETTER S
⟨U015B⟩ ⟨s⟩;⟨ACUTE⟩;⟨SMALL⟩;IGNORE % LATIN SMALL LETTER S WITH ACUTE
⟨U015A⟩ ⟨s⟩;⟨ACUTE⟩;⟨CAPITAL⟩;IGNORE % LATIN CAPITAL LETTER S WITH ACUTE
⟨U015D⟩ ⟨s⟩;⟨CIRCUMFLEX⟩;⟨SMALL⟩;IGNORE % LATIN SMALL LETTER S WITH CIRCUMFLEX
⟨U015C⟩ ⟨s⟩;⟨CIRCUMFLEX⟩;⟨CAPITAL⟩;IGNORE % LATIN CAPITAL LETTER S WITH CIRCUMFLEX
⟨U0161⟩ ⟨s⟩;⟨CARON⟩;⟨SMALL⟩;IGNORE % LATIN SMALL LETTER S WITH CARON
⟨U0160⟩ ⟨s⟩;⟨CARON⟩;⟨CAPITAL⟩;IGNORE % LATIN CAPITAL LETTER S WITH CARON
⟨U1E61⟩ ⟨s⟩;⟨DOT_ABOVE⟩;⟨SMALL⟩;IGNORE % LATIN SMALL LETTER S WITH DOT ABOVE
⟨U1E60⟩ ⟨s⟩;⟨DOT_ABOVE⟩;⟨CAPITAL⟩;IGNORE % LATIN CAPITAL LETTER S WITH DOT ABOVE
⟨U015F⟩ ⟨s⟩;⟨CEDILLA⟩;⟨SMALL⟩;IGNORE % LATIN SMALL LETTER S WITH CEDILLA
⟨U015E⟩ ⟨s⟩;⟨CEDILLA⟩;⟨CAPITAL⟩;IGNORE % LATIN CAPITAL LETTER S WITH CEDILLA
⟨U017F⟩ ⟨s⟩;⟨MODIFIED⟩;⟨SMALL⟩;IGNORE % LATIN SMALL LETTER LONG S
⟨U00DF⟩ ″⟨s⟩⟨s⟩″;″⟨MODIFIED⟩⟨BLANK⟩″;″⟨SMALL⟩⟨SMALL⟩″;IGNORE % LATIN SMALL LETTER SHARP S
⟨U0074⟩ ⟨t⟩;⟨BLANK⟩;⟨SMALL⟩;IGNORE % LATIN SMALL LETTER T
⟨U0054⟩ ⟨t⟩;⟨BLANK⟩;⟨CAPITAL⟩;IGNORE % LATIN CAPITAL LETTER T
⟨U0165⟩ ⟨t⟩;⟨CARON⟩;⟨SMALL⟩;IGNORE % LATIN SMALL LETTER T WITH CARON
⟨U0164⟩ ⟨t⟩;⟨CARON⟩;⟨CAPITAL⟩;IGNORE % LATIN CAPITAL LETTER T WITH CARON
⟨U1E6B⟩ ⟨t⟩;⟨DOT_ABOVE⟩;⟨SMALL⟩;IGNORE % LATIN SMALL LETTER T WITH DOT ABOVE
⟨U1E6A⟩ ⟨t⟩;⟨DOT_ABOVE⟩;⟨CAPITAL⟩;IGNORE % LATIN CAPITAL LETTER T WITH DOT ABOVE
⟨U0163⟩ ⟨t⟩;⟨CEDILLA⟩;⟨SMALL⟩;IGNORE % LATIN SMALL LETTER T WITH CEDILLA
⟨U0162⟩ ⟨t⟩;⟨CEDILLA⟩;⟨CAPITAL⟩;IGNORE % LATIN CAPITAL LETTER T WITH CEDILLA
⟨U0167⟩ ⟨t⟩;⟨STROKE⟩;⟨SMALL⟩;IGNORE % LATIN SMALL LETTER T WITH STROKE

⟨U0166⟩ ⟨t⟩;⟨STROKE⟩;⟨CAPITAL⟩;IGNORE % LATIN CAPITAL LETTER T WITH STROKE
⟨U0075⟩ ⟨u⟩;⟨BLANK⟩;⟨SMALL⟩;IGNORE % LATIN SMALL LETTER U
⟨U0055⟩ ⟨u⟩;⟨BLANK⟩;⟨CAPITAL⟩;IGNORE % LATIN CAPITAL LETTER U
⟨U00FA⟩ ⟨u⟩;⟨ACUTE⟩;⟨SMALL⟩;IGNORE % LATIN SMALL LETTER U WITH ACUTE
⟨U00DA⟩ ⟨u⟩;⟨ACUTE⟩;⟨CAPITAL⟩;IGNORE % LATIN CAPITAL LETTER U WITH ACUTE
⟨U00F9⟩ ⟨u⟩;⟨GRAVE⟩;⟨SMALL⟩;IGNORE % LATIN SMALL LETTER U WITH GRAVE
⟨U00D9⟩ ⟨u⟩;⟨GRAVE⟩;⟨CAPITAL⟩;IGNORE % LATIN CAPITAL LETTER U WITH GRAVE
⟨U016D⟩ ⟨u⟩;⟨BREVE⟩;⟨SMALL⟩;IGNORE % LATIN SMALL LETTER U WITH BREVE
⟨U016C⟩ ⟨u⟩;⟨BREVE⟩;⟨CAPITAL⟩;IGNORE % LATIN CAPITAL LETTER U WITH BREVE
⟨U00FB⟩ ⟨u⟩;⟨CIRCUMFLEX⟩;⟨SMALL⟩;IGNORE % LATIN SMALL LETTER U WITH CIRCUMFLEX
⟨U00DB⟩ ⟨u⟩;⟨CIRCUMFLEX⟩;⟨CAPITAL⟩;IGNORE % LATIN CAPITAL LETTER U WITH CIRCUMFLEX
⟨U016F⟩ ⟨u⟩;⟨RING_ABOVE⟩;⟨SMALL⟩;IGNORE % LATIN SMALL LETTER U WITH RING ABOVE
⟨U016E⟩ ⟨u⟩;⟨RING_ABOVE⟩;⟨CAPITAL⟩;IGNORE % LATIN CAPITAL LETTER U WITH RING ABOVE
⟨U00FC⟩ ⟨u⟩;⟨DIAERESIS⟩;⟨SMALL⟩;IGNORE % LATIN SMALL LETTER U WITH DIAERESIS
⟨U00DC⟩ ⟨u⟩;⟨DIAERESIS⟩;⟨CAPITAL⟩;IGNORE % LATIN CAPITAL LETTER U WITH DIAERESIS
⟨U0171⟩ ⟨u⟩;⟨DOUBLE_ACUTE⟩;⟨SMALL⟩;IGNORE % LATIN SMALL LETTER U WITH DOUBLE ACUTE
⟨U0170⟩ ⟨u⟩;⟨DOUBLE_ACUTE⟩;⟨CAPITAL⟩;IGNORE % LATIN CAPITAL LETTER U WITH DOUBLE ACUTE
⟨U0169⟩ ⟨u⟩;⟨TILDE⟩;⟨SMALL⟩;IGNORE % LATIN SMALL LETTER U WITH TILDE
⟨U0168⟩ ⟨u⟩;⟨TILDE⟩;⟨CAPITAL⟩;IGNORE % LATIN CAPITAL LETTER U WITH TILDE
⟨U0173⟩ ⟨u⟩;⟨OGONEK⟩;⟨SMALL⟩;IGNORE % LATIN SMALL LETTER U WITH OGONEK
⟨U0172⟩ ⟨u⟩;⟨OGONEK⟩;⟨CAPITAL⟩;IGNORE % LATIN CAPITAL LETTER U WITH OGONEK
⟨U016B⟩ ⟨u⟩;⟨MACRON⟩;⟨SMALL⟩;IGNORE % LATIN SMALL LETTER U WITH MACRON
⟨U016A⟩ ⟨u⟩;⟨MACRON⟩;⟨CAPITAL⟩;IGNORE % LATIN CAPITAL LETTER U WITH MACRON
⟨U0076⟩ ⟨v⟩;⟨BLANK⟩;⟨SMALL⟩;IGNORE % LATIN SMALL LETTER V
⟨U0056⟩ ⟨v⟩;⟨BLANK⟩;⟨CAPITAL⟩;IGNORE % LATIN CAPITAL LETTER V
⟨U0077⟩ ⟨w⟩;⟨BLANK⟩;⟨SMALL⟩;IGNORE % LATIN SMALL LETTER W
⟨U0057⟩ ⟨w⟩;⟨BLANK⟩;⟨CAPITAL⟩;IGNORE % LATIN CAPITAL LETTER W
⟨U1E83⟩ ⟨w⟩;⟨ACUTE⟩;⟨SMALL⟩;IGNORE % LATIN SMALL LETTER W WITH ACUTE
⟨U1E82⟩ ⟨w⟩;⟨ACUTE⟩;⟨CAPITAL⟩;IGNORE % LATIN CAPITAL LETTER W WITH ACUTE
⟨U1E81⟩ ⟨w⟩;⟨GRAVE⟩;⟨SMALL⟩;IGNORE % LATIN SMALL LETTER W WITH GRAVE
⟨U1E80⟩ ⟨w⟩;⟨GRAVE⟩;⟨CAPITAL⟩;IGNORE % LATIN CAPITAL LETTER W WITH GRAVE
⟨U0175⟩ ⟨w⟩;⟨CIRCUMFLEX⟩;⟨SMALL⟩;IGNORE % LATIN SMALL LETTER W WITH CIRCUMFLEX
⟨U0174⟩ ⟨w⟩;⟨CIRCUMFLEX⟩;⟨CAPITAL⟩;IGNORE % LATIN CAPITAL LETTER W WITH CIRCUMFLEX
⟨U1E85⟩ ⟨w⟩;⟨DIAERESIS⟩;⟨SMALL⟩;IGNORE % LATIN SMALL LETTER W WITH DIAERESIS
⟨U1E84⟩ ⟨w⟩;⟨DIAERESIS⟩;⟨CAPITAL⟩;IGNORE % LATIN CAPITAL LETTER W WITH DIAERESIS
⟨U0078⟩ ⟨x⟩;⟨BLANK⟩;⟨SMALL⟩;IGNORE % LATIN SMALL LETTER X
⟨U0058⟩ ⟨x⟩;⟨BLANK⟩;⟨CAPITAL⟩;IGNORE % LATIN CAPITAL LETTER X
⟨U0079⟩ ⟨y⟩;⟨BLANK⟩;⟨SMALL⟩;IGNORE % LATIN SMALL LETTER Y
⟨U0059⟩ ⟨y⟩;⟨BLANK⟩;⟨CAPITAL⟩;IGNORE % LATIN CAPITAL LETTER Y
⟨U00FD⟩ ⟨y⟩;⟨ACUTE⟩;⟨SMALL⟩;IGNORE % LATIN SMALL LETTER Y WITH ACUTE
⟨U00DD⟩ ⟨y⟩;⟨ACUTE⟩;⟨CAPITAL⟩;IGNORE % LATIN CAPITAL LETTER Y WITH ACUTE
⟨U1EF3⟩ ⟨y⟩;⟨GRAVE⟩;⟨SMALL⟩;IGNORE % LATIN SMALL LETTER Y WITH GRAVE
⟨U1EF2⟩ ⟨y⟩;⟨GRAVE⟩;⟨CAPITAL⟩;IGNORE % LATIN CAPITAL LETTER Y WITH GRAVE
⟨U0177⟩ ⟨y⟩;⟨CIRCUMFLEX⟩;⟨SMALL⟩;IGNORE % LATIN SMALL LETTER Y WITH CIRCUMFLEX
⟨U0176⟩ ⟨y⟩;⟨CIRCUMFLEX⟩;⟨CAPITAL⟩;IGNORE % LATIN CAPITAL LETTER Y WITH CIRCUMFLEX
⟨U00FF⟩ ⟨y⟩;⟨DIAERESIS⟩;⟨SMALL⟩;IGNORE % LATIN SMALL LETTER Y WITH DIAERESIS
⟨U0178⟩ ⟨y⟩;⟨DIAERESIS⟩;⟨CAPITAL⟩;IGNORE % LATIN CAPITAL LETTER Y WITH DIAERESIS
⟨U007A⟩ ⟨z⟩;⟨BLANK⟩;⟨SMALL⟩;IGNORE % LATIN SMALL LETTER Z
⟨U005A⟩ ⟨z⟩;⟨BLANK⟩;⟨CAPITAL⟩;IGNORE % LATIN CAPITAL LETTER Z
⟨U017A⟩ ⟨z⟩;⟨ACUTE⟩;⟨SMALL⟩;IGNORE % LATIN SMALL LETTER Z WITH ACUTE
⟨U0179⟩ ⟨z⟩;⟨ACUTE⟩;⟨CAPITAL⟩;IGNORE % LATIN CAPITAL LETTER Z WITH ACUTE
⟨U017E⟩ ⟨z⟩;⟨CARON⟩;⟨SMALL⟩;IGNORE % LATIN SMALL LETTER Z WITH CARON
⟨U017D⟩ ⟨z⟩;⟨CARON⟩;⟨CAPITAL⟩;IGNORE % LATIN CAPITAL LETTER Z WITH CARON
⟨U017C⟩ ⟨z⟩;⟨DOT_ABOVE⟩;⟨SMALL⟩;IGNORE % LATIN SMALL LETTER Z WITH DOT ABOVE
⟨U017B⟩ ⟨z⟩;⟨DOT_ABOVE⟩;⟨CAPITAL⟩;IGNORE % LATIN CAPITAL LETTER Z WITH DOT ABOVE
⟨U0292⟩ ⟨z⟩;⟨MODIFIED⟩;⟨SMALL⟩;IGNORE % LATIN SMALL LETTER EZH
⟨U01B7⟩ ⟨z⟩;⟨MODIFIED⟩;⟨CAPITAL⟩;IGNORE % LATIN CAPITAL LETTER EZH

⟨U01EF⟩ ⟨z⟩;⟨MODIFIED+CARON⟩;⟨SMALL⟩;IGNORE % LATIN SMALL LETTER EZH WITH CARON
⟨U01EE⟩ ⟨z⟩;⟨MODIFIED+CARON⟩;⟨CAPITAL⟩;IGNORE % LATIN CAPITAL LETTER EZH WITH CARON
⟨U00FE⟩ ⟨thorn⟩;⟨BLANK⟩;⟨SMALL⟩;IGNORE % LATIN SMALL LETTER THORN
⟨U00DE⟩ ⟨thorn⟩;⟨BLANK⟩;⟨CAPITAL⟩;IGNORE % LATIN CAPITAL LETTER THORN

section EORGreek
⟨U03B1⟩ ⟨alpha⟩;⟨BLANK⟩;⟨SMALL⟩;IGNORE % GREEK SMALL LETTER ALPHA
⟨U0391⟩ ⟨alpha⟩;⟨BLANK⟩;⟨CAPITAL⟩;IGNORE % GREEK CAPITAL LETTER ALPHA
⟨U1F00⟩ ⟨alpha⟩;⟨PSILI⟩;⟨SMALL⟩;IGNORE % GREEK SMALL LETTER ALPHA WITH PSILI
⟨U1F08⟩ ⟨alpha⟩;⟨PSILI⟩;⟨CAPITAL⟩;IGNORE % GREEK CAPITAL LETTER ALPHA WITH PSILI
⟨U1F04⟩ ⟨alpha⟩;⟨PSILI+OXIA⟩;⟨SMALL⟩;IGNORE % GREEK SMALL LETTER ALPHA WITH PSILI AND OXIA
⟨U1F0C⟩ ⟨alpha⟩;⟨PSILI+OXIA⟩;⟨CAPITAL⟩;IGNORE % GREEK CAPITAL LETTER ALPHA WITH PSILI AND OXIA
⟨U1F8C⟩ ⟨alpha⟩;⟨PSILI+OXIA+PROSGEGRAMMENI⟩;⟨CAPITAL⟩;IGNORE % GREEK CAPITAL LETTER ALPHA WITH PSILI AND OXIA AND PROSGEGRAMMENI
⟨U1F84⟩ ⟨alpha⟩;⟨PSILI+OXIA+YPOGEGRAMMENI⟩;⟨SMALL⟩;IGNORE % GREEK SMALL LETTER ALPHA WITH PSILI AND OXIA AND YPOGEGRAMMENI
⟨U1F02⟩ ⟨alpha⟩;⟨PSILI+VARIA⟩;⟨SMALL⟩;IGNORE % GREEK SMALL LETTER ALPHA WITH PSILI AND VARIA
⟨U1F0A⟩ ⟨alpha⟩;⟨PSILI+VARIA⟩;⟨CAPITAL⟩;IGNORE % GREEK CAPITAL LETTER ALPHA WITH PSILI AND VARIA
⟨U1F8A⟩ ⟨alpha⟩;⟨PSILI+VARIA+PROSGEGRAMMENI⟩;⟨CAPITAL⟩;IGNORE % GREEK CAPITAL LETTER ALPHA WITH PSILI AND VARIA AND PROSGEGRAMMENI
⟨U1F82⟩ ⟨alpha⟩;⟨PSILI+VARIA+YPOGEGRAMMENI⟩;⟨SMALL⟩;IGNORE % GREEK SMALL LETTER ALPHA WITH PSILI AND VARIA AND YPOGEGRAMMENI
⟨U1F8E⟩ ⟨alpha⟩;⟨PSILI+PERISPOMENI+PROSGEGRAMMENI⟩;⟨CAPITAL⟩;IGNORE % GREEK CAPITAL LETTER ALPHA WITH PSILI AND PERISPOMENI AND PROSGEGRAMMENI
⟨U1F86⟩ ⟨alpha⟩;⟨PSILI+PERISPOMENI+YPOGEGRAMMENI⟩;⟨SMALL⟩;IGNORE % GREEK SMALL LETTER ALPHA WITH PSILI AND PERISPOMENI AND YPOGEGRAMMENI
⟨U1F06⟩ ⟨alpha⟩;⟨PSILI+PERISPOMENI⟩;⟨SMALL⟩;IGNORE % GREEK SMALL LETTER ALPHA WITH PSILI AND PERISPOMENI
⟨U1F0E⟩ ⟨alpha⟩;⟨PSILI+PERISPOMENI⟩;⟨CAPITAL⟩;IGNORE % GREEK CAPITAL LETTER ALPHA WITH PSILI AND PERISPOMENI
⟨U1F88⟩ ⟨alpha⟩;⟨PSILI+PROSGEGRAMMENI⟩;⟨CAPITAL⟩;IGNORE % GREEK CAPITAL LETTER ALPHA WITH PSILI AND PROSGEGRAMMENI
⟨U1F80⟩ ⟨alpha⟩;⟨PSILI+YPOGEGRAMMENI⟩;⟨SMALL⟩;IGNORE % GREEK SMALL LETTER ALPHA WITH PSILI AND YPOGEGRAMMENI
⟨U1F01⟩ ⟨alpha⟩;⟨DASIA⟩;⟨SMALL⟩;IGNORE % GREEK SMALL LETTER ALPHA WITH DASIA
⟨U1F09⟩ ⟨alpha⟩;⟨DASIA⟩;⟨CAPITAL⟩;IGNORE % GREEK CAPITAL LETTER ALPHA WITH DASIA
⟨U1F05⟩ ⟨alpha⟩;⟨DASIA+OXIA⟩;⟨SMALL⟩;IGNORE % GREEK SMALL LETTER ALPHA WITH DASIA AND OXIA
⟨U1F0D⟩ ⟨alpha⟩;⟨DASIA+OXIA⟩;⟨CAPITAL⟩;IGNORE % GREEK CAPITAL LETTER ALPHA WITH DASIA AND OXIA
⟨U1F8D⟩ ⟨alpha⟩;⟨DASIA+OXIA+PROSGEGRAMMENI⟩;⟨CAPITAL⟩;IGNORE % GREEK CAPITAL LETTER ALPHA WITH DASIA AND OXIA AND PROSGEGRAMMENI
⟨U1F85⟩ ⟨alpha⟩;⟨DASIA+OXIA+YPOGEGRAMMENI⟩;⟨SMALL⟩;IGNORE % GREEK SMALL LETTER ALPHA WITH DASIA AND OXIA AND YPOGEGRAMMENI
⟨U1F03⟩ ⟨alpha⟩;⟨DASIA+VARIA⟩;⟨SMALL⟩;IGNORE % GREEK SMALL LETTER ALPHA WITH DASIA AND VARIA
⟨U1F0B⟩ ⟨alpha⟩;⟨DASIA+VARIA⟩;⟨CAPITAL⟩;IGNORE % GREEK CAPITAL LETTER ALPHA WITH DASIA AND VARIA
⟨U1F8B⟩ ⟨alpha⟩;⟨DASIA+VARIA+PROSGEGRAMMENI⟩;⟨CAPITAL⟩;IGNORE % GREEK CAPITAL LETTER ALPHA WITH DASIA AND VARIA AND PROSGEGRAMMENI
⟨U1F83⟩ ⟨alpha⟩;⟨DASIA+VARIA+YPOGEGRAMMENI⟩;⟨SMALL⟩;IGNORE % GREEK SMALL LETTER ALPHA WITH DASIA AND VARIA AND YPOGEGRAMMENI
⟨U1F07⟩ ⟨alpha⟩;⟨DASIA+PERISPOMENI⟩;⟨SMALL⟩;IGNORE % GREEK SMALL LETTER ALPHA WITH DASIA AND PERISPOMENI

⟨U1F0F⟩ ⟨alpha⟩;⟨DASIA+PERISPOMENI⟩;⟨CAPITAL⟩;IGNORE % GREEK CAPITAL LETTER ALPHA WITH DASIA AND PERISPOMENI
⟨U1F8F⟩ ⟨alpha⟩;⟨DASIA+PROSGEGRAMMENI⟩;⟨CAPITAL⟩;IGNORE % GREEK CAPITAL LETTER ALPHA WITH DASIA AND PERISPOMENI AND PROSGEGRAMMENI
⟨U1F89⟩ ⟨alpha⟩;⟨DASIA+PROSGEGRAMMENI⟩;⟨CAPITAL⟩;IGNORE % GREEK CAPITAL LETTER ALPHA WITH DASIA AND PROSGEGRAMMENI
⟨U1F87⟩ ⟨alpha⟩;⟨DASIA+YPOGEGRAMMENI⟩;⟨SMALL⟩;IGNORE % GREEK SMALL LETTER ALPHA WITH DASIA AND PERISPOMENI AND YPOGEGRAMMENI
⟨U1F81⟩ ⟨alpha⟩;⟨DASIA+YPOGEGRAMMENI⟩;⟨SMALL⟩;IGNORE % GREEK SMALL LETTER ALPHA WITH DASIA AND YPOGEGRAMMENI
⟨U1F71⟩ ⟨alpha⟩;⟨ACUTE⟩;⟨SMALL⟩;IGNORE % GREEK SMALL LETTER ALPHA WITH OXIA
⟨U1FBB⟩ ⟨alpha⟩;⟨ACUTE⟩;⟨CAPITAL⟩;IGNORE % GREEK CAPITAL LETTER ALPHA WITH OXIA
⟨U1FB4⟩ ⟨alpha⟩;⟨OXIA+YPOGEGRAMMENI⟩;⟨SMALL⟩;IGNORE % GREEK SMALL LETTER ALPHA WITH OXIA AND YPOGEGRAMMENI
⟨U1F70⟩ ⟨alpha⟩;⟨GRAVE⟩;⟨SMALL⟩;IGNORE % GREEK SMALL LETTER ALPHA WITH VARIA
⟨U1FBA⟩ ⟨alpha⟩;⟨GRAVE⟩;⟨CAPITAL⟩;IGNORE % GREEK CAPITAL LETTER ALPHA WITH VARIA
⟨U1FB2⟩ ⟨alpha⟩;⟨VARIA+YPOGEGRAMMENI⟩;⟨SMALL⟩;IGNORE % GREEK SMALL LETTER ALPHA WITH VARIA AND YPOGEGRAMMENI
⟨U1FB6⟩ ⟨alpha⟩;⟨PERISPOMENI⟩;⟨SMALL⟩;IGNORE % GREEK SMALL LETTER ALPHA WITH PERISPOMENI
⟨U1FB7⟩ ⟨alpha⟩;⟨PERISPOMENI+YPOGEGRAMMENI⟩;⟨SMALL⟩;IGNORE % GREEK SMALL LETTER ALPHA WITH PERISPOMENI AND YPOGRAMMENI
⟨U03AC⟩ ⟨alpha⟩;⟨TONOS⟩;⟨SMALL⟩;IGNORE % GREEK SMALL LETTER ALPHA WITH TONOS
⟨U0386⟩ ⟨alpha⟩;⟨TONOS⟩;⟨CAPITAL⟩;IGNORE % GREEK CAPITAL LETTER ALPHA WITH TONOS
⟨U1FBC⟩ ⟨alpha⟩;⟨PROSGEGRAMMENI⟩;⟨CAPITAL⟩;IGNORE % GREEK CAPITAL LETTER ALPHA WITH PROSGEGRAMMENI
⟨U1FB3⟩ ⟨alpha⟩;⟨YPOGEGRAMMENI⟩;⟨SMALL⟩;IGNORE % GREEK SMALL LETTER ALPHA WITH YPOGEGRAMMENI
⟨U1FB0⟩ ⟨alpha⟩;⟨BREVE⟩;⟨SMALL⟩;IGNORE % GREEK SMALL LETTER ALPHA WITH VRACHY
⟨U1FB8⟩ ⟨alpha⟩;⟨BREVE⟩;⟨CAPITAL⟩;IGNORE % GREEK CAPITAL LETTER ALPHA WITH VRACHY
⟨U1FB1⟩ ⟨alpha⟩;⟨MACRON⟩;⟨SMALL⟩;IGNORE % GREEK SMALL LETTER ALPHA WITH MACRON
⟨U1FB9⟩ ⟨alpha⟩;⟨MACRON⟩;⟨CAPITAL⟩;IGNORE % GREEK CAPITAL LETTER ALPHA WITH MACRON
⟨U03B2⟩ ⟨beta⟩;⟨BLANK⟩;⟨SMALL⟩;IGNORE % GREEK SMALL LETTER BETA
⟨U0392⟩ ⟨beta⟩;⟨BLANK⟩;⟨CAPITAL⟩;IGNORE % GREEK CAPITAL LETTER BETA
⟨U03B3⟩ ⟨gamma⟩;⟨BLANK⟩;⟨SMALL⟩;IGNORE % GREEK SMALL LETTER GAMMA
⟨U0393⟩ ⟨gamma⟩;⟨BLANK⟩;⟨CAPITAL⟩;IGNORE % GREEK CAPITAL LETTER GAMMA
⟨U03B4⟩ ⟨delta⟩;⟨BLANK⟩;⟨SMALL⟩;IGNORE % GREEK SMALL LETTER DELTA
⟨U0394⟩ ⟨delta⟩;⟨BLANK⟩;⟨CAPITAL⟩;IGNORE % GREEK CAPITAL LETTER DELTA
⟨U03B5⟩ ⟨epsilon⟩;⟨BLANK⟩;⟨SMALL⟩;IGNORE % GREEK SMALL LETTER EPSILON
⟨U0395⟩ ⟨epsilon⟩;⟨BLANK⟩;⟨CAPITAL⟩;IGNORE % GREEK CAPITAL LETTER EPSILON
⟨U1F10⟩ ⟨epsilon⟩;⟨PSILI⟩;⟨SMALL⟩;IGNORE % GREEK SMALL LETTER EPSILON WITH PSILI
⟨U1F18⟩ ⟨epsilon⟩;⟨PSILI⟩;⟨CAPITAL⟩;IGNORE % GREEK CAPITAL LETTER EPSILON WITH PSILI
⟨U1F14⟩ ⟨epsilon⟩;⟨PSILI+OXIA⟩;⟨SMALL⟩;IGNORE % GREEK SMALL LETTER EPSILON WITH PSILI AND OXIA
⟨U1F1C⟩ ⟨epsilon⟩;⟨PSILI+OXIA⟩;⟨CAPITAL⟩;IGNORE % GREEK CAPITAL LETTER EPSILON WITH PSILI AND OXIA
⟨U1F12⟩ ⟨epsilon⟩;⟨PSILI+VARIA⟩;⟨SMALL⟩;IGNORE % GREEK SMALL LETTER EPSILON WITH PSILI AND VARIA
⟨U1F1A⟩ ⟨epsilon⟩;⟨PSILI+VARIA⟩;⟨CAPITAL⟩;IGNORE % GREEK CAPITAL LETTER EPSILON WITH PSILI AND VARIA
⟨U1F11⟩ ⟨epsilon⟩;⟨DASIA⟩;⟨SMALL⟩;IGNORE % GREEK SMALL LETTER EPSILON WITH DASIA
⟨U1F19⟩ ⟨epsilon⟩;⟨DASIA⟩;⟨CAPITAL⟩;IGNORE % GREEK CAPITAL LETTER EPSILON WITH DASIA
⟨U1F15⟩ ⟨epsilon⟩;⟨DASIA+OXIA⟩;⟨SMALL⟩;IGNORE % GREEK SMALL LETTER EPSILON WITH DASIA AND OXIA
⟨U1F1D⟩ ⟨epsilon⟩;⟨DASIA+OXIA⟩;⟨CAPITAL⟩;IGNORE % GREEK CAPITAL LETTER EPSILON WITH DASIA AND OXIA

⟨U1F13⟩ ⟨epsilon⟩;⟨DASIA+VARIA⟩;⟨SMALL⟩;IGNORE % GREEK SMALL LETTER EPSILON WITH DASIA AND VARIA
⟨U1F1B⟩ ⟨epsilon⟩;⟨DASIA+VARIA⟩;⟨CAPITAL⟩;IGNORE % GREEK CAPITAL LETTER EPSILON WITH DASIA AND VARIA
⟨U1F73⟩ ⟨epsilon⟩;⟨ACUTE⟩;⟨SMALL⟩;IGNORE % GREEK SMALL LETTER EPSILON WITH OXIA
⟨U1FC9⟩ ⟨epsilon⟩;⟨ACUTE⟩;⟨CAPITAL⟩;IGNORE % GREEK CAPITAL LETTER EPSILON WITH OXIA
⟨U1F72⟩ ⟨epsilon⟩;⟨GRAVE⟩;⟨SMALL⟩;IGNORE % GREEK SMALL LETTER EPSILON WITH VARIA
⟨U1FC8⟩ ⟨epsilon⟩;⟨GRAVE⟩;⟨CAPITAL⟩;IGNORE % GREEK CAPITAL LETTER EPSILON WITH VARIA
⟨U03AD⟩ ⟨epsilon⟩;⟨TONOS⟩;⟨SMALL⟩;IGNORE % GREEK SMALL LETTER EPSILON WITH TONOS
⟨U0388⟩ ⟨epsilon⟩;⟨TONOS⟩;⟨CAPITAL⟩;IGNORE % GREEK CAPITAL LETTER EPSILON WITH TONOS
⟨U03B6⟩ ⟨zeta⟩;⟨BLANK⟩;⟨SMALL⟩;IGNORE % GREEK SMALL LETTER ZETA
⟨U0396⟩ ⟨zeta⟩;⟨BLANK⟩;⟨CAPITAL⟩;IGNORE % GREEK CAPITAL LETTER ZETA
⟨U03B7⟩ ⟨eta⟩;⟨BLANK⟩;⟨SMALL⟩;IGNORE % GREEK SMALL LETTER ETA
⟨U0397⟩ ⟨eta⟩;⟨BLANK⟩;⟨CAPITAL⟩;IGNORE % GREEK CAPITAL LETTER ETA
⟨U1F20⟩ ⟨eta⟩;⟨PSILI⟩;⟨SMALL⟩;IGNORE % GREEK SMALL LETTER ETA WITH PSILI
⟨U1F28⟩ ⟨eta⟩;⟨PSILI⟩;⟨CAPITAL⟩;IGNORE % GREEK CAPITAL LETTER ETA WITH PSILI
⟨U1F24⟩ ⟨eta⟩;⟨PSILI+OXIA⟩;⟨SMALL⟩;IGNORE % GREEK SMALL LETTER ETA WITH PSILI AND OXIA
⟨U1F2C⟩ ⟨eta⟩;⟨PSILI+OXIA⟩;⟨CAPITAL⟩;IGNORE % GREEK CAPITAL LETTER ETA WITH PSILI AND OXIA
⟨U1F9C⟩ ⟨eta⟩;⟨PSILI+OXIA+PROSGEGRAMMENI⟩;⟨CAPITAL⟩;IGNORE % GREEK CAPITAL LETTER ETA WITH PSILI AND OXIA AND PROSGEGRAMMENI
⟨U1F94⟩ ⟨eta⟩;⟨PSILI+OXIA+YPOGEGRAMMENI⟩;⟨SMALL⟩;IGNORE % GREEK SMALL LETTER ETA WITH PSILI AND OXIA AND YPOGEGRAMMENI
⟨U1F22⟩ ⟨eta⟩;⟨PSILI+VARIA⟩;⟨SMALL⟩;IGNORE % GREEK SMALL LETTER ETA WITH PSILI AND VARIA
⟨U1F2A⟩ ⟨eta⟩;⟨PSILI+VARIA⟩;⟨CAPITAL⟩;IGNORE % GREEK CAPITAL LETTER ETA WITH PSILI AND VARIA
⟨U1F9A⟩ ⟨eta⟩;⟨PSILI+VARIA+PROSGEGRAMMENI⟩;⟨CAPITAL⟩;IGNORE % GREEK CAPITAL LETTER ETA WITH PSILI AND VARIA AND PROSGEGRAMMENI
⟨U1F92⟩ ⟨eta⟩;⟨PSILI+VARIA+YPOGEGRAMMENI⟩;⟨SMALL⟩;IGNORE % GREEK SMALL LETTER ETA WITH PSILI AND VARIA AND YPOGEGRAMMENI
⟨U1F9E⟩ ⟨eta⟩;⟨PSILI+PERISPOMENI+PROSGEGRAMMENI⟩;⟨CAPITAL⟩;IGNORE % GREEK CAPITAL LETTER ETA WITH PSILI AND PERISPOMENI AND PROSGEGRAMMENI
⟨U1F96⟩ ⟨eta⟩;⟨PSILI+PERISPOMENI+YPOGEGRAMMENI⟩;⟨SMALL⟩;IGNORE % GREEK SMALL LETTER ETA WITH PSILI AND PERISPOMENI AND YPOGEGRAMMENI
⟨U1F26⟩ ⟨eta⟩;⟨PSILI+PERISPOMENI⟩;⟨SMALL⟩;IGNORE % GREEK SMALL LETTER ETA WITH PSILI AND PERISPOMENI
⟨U1F2E⟩ ⟨eta⟩;⟨PSILI+PERISPOMENI⟩;⟨CAPITAL⟩;IGNORE % GREEK CAPITAL LETTER ETA WITH PSILI AND PERISPOMENI
⟨U1F98⟩ ⟨eta⟩;⟨PSILI+PROSGEGRAMMENI⟩;⟨CAPITAL⟩;IGNORE % GREEK CAPITAL LETTER ETA WITH PSILI AND PROSGEGRAMMENI
⟨U1F90⟩ ⟨eta⟩;⟨PSILI+YPOGEGRAMMENI⟩;⟨SMALL⟩;IGNORE % GREEK SMALL LETTER ETA WITH PSILI AND YPOGEGRAMMENI
⟨U1F21⟩ ⟨eta⟩;⟨DASIA⟩;⟨SMALL⟩;IGNORE % GREEK SMALL LETTER ETA WITH DASIA
⟨U1F29⟩ ⟨eta⟩;⟨DASIA⟩;⟨CAPITAL⟩;IGNORE % GREEK CAPITAL LETTER ETA WITH DASIA
⟨U1F25⟩ ⟨eta⟩;⟨DASIA+OXIA⟩;⟨SMALL⟩;IGNORE % GREEK SMALL LETTER ETA WITH DASIA AND OXIA
⟨U1F2D⟩ ⟨eta⟩;⟨DASIA+OXIA⟩;⟨CAPITAL⟩;IGNORE % GREEK CAPITAL LETTER ETA WITH DASIA AND OXIA
⟨U1F9D⟩ ⟨eta⟩;⟨DASIA+OXIA+PROSGEGRAMMENI⟩;⟨CAPITAL⟩;IGNORE % GREEK CAPITAL LETTER ETA WITH DASIA AND OXIA AND PROSGEGRAMMENI
⟨U1F95⟩ ⟨eta⟩;⟨DASIA+OXIA+YPOGEGRAMMENI⟩;⟨SMALL⟩;IGNORE % GREEK SMALL LETTER ETA WITH DASIA AND OXIA AND YPOGEGRAMMENI
⟨U1F23⟩ ⟨eta⟩;⟨DASIA+VARIA⟩;⟨SMALL⟩;IGNORE % GREEK SMALL LETTER ETA WITH DASIA AND VARIA

15

⟨U1F2B⟩ ⟨eta⟩;⟨DASIA+VARIA⟩;⟨CAPITAL⟩;IGNORE % GREEK CAPITAL LETTER ETA WITH DASIA AND VARIA

⟨U1F9B⟩ ⟨eta⟩;⟨DASIA+VARIA+PROSGEGRAMMENI⟩;⟨CAPITAL⟩;IGNORE % GREEK CAPITAL LETTER ETA WITH DASIA AND VARIA AND PROSGEGRAMMENI

⟨U1F93⟩ ⟨eta⟩;⟨DASIA+VARIA+YPOGEGRAMMENI⟩;⟨SMALL⟩;IGNORE % GREEK SMALL LETTER ETA WITH DASIA AND VARIA AND YPOGEGRAMMENI

⟨U1F27⟩ ⟨eta⟩;⟨DASIA+PERISPOMENI⟩;⟨SMALL⟩;IGNORE % GREEK SMALL LETTER ETA WITH DASIA AND PERISPOMENI

⟨U1F2F⟩ ⟨eta⟩;⟨DASIA+PERISPOMENI⟩;⟨CAPITAL⟩;IGNORE % GREEK CAPITAL LETTER ETA WITH DASIA AND PERISPOMENI

⟨U1F9F⟩ ⟨eta⟩;⟨DASIA+PERISPOMENI+PROSGEGRAMMENI⟩;⟨CAPITAL⟩;IGNORE % GREEK CAPITAL LETTER ETA WITH DASIA AND PERISPOMENI AND PROSGEGRAMMENI

⟨U1F97⟩ ⟨eta⟩;⟨DASIA+PERISPOMENI+YPOGEGRAMMENI⟩;⟨SMALL⟩;IGNORE % GREEK SMALL LETTER ETA WITH DASIA AND PERISPOMENI AND YPOGEGRAMMENI

⟨U1F99⟩ ⟨eta⟩;⟨DASIA+PROSGEGRAMMENI⟩;⟨CAPITAL⟩;IGNORE % GREEK CAPITAL LETTER ETA WITH DASIA AND PROSGEGRAMMENI

⟨U1F91⟩ ⟨eta⟩;⟨DASIA+YPOGEGRAMMENI⟩;⟨SMALL⟩;IGNORE % GREEK SMALL LETTER ETA WITH DASIA AND YPOGEGRAMMENI

⟨U1F75⟩ ⟨eta⟩;⟨ACUTE⟩;⟨SMALL⟩;IGNORE % GREEK SMALL LETTER ETA WITH OXIA

⟨U1FCB⟩ ⟨eta⟩;⟨ACUTE⟩;⟨CAPITAL⟩;IGNORE % GREEK CAPITAL LETTER ETA WITH OXIA

⟨U1FC4⟩ ⟨eta⟩;⟨OXIA+YPOGEGRAMMENI⟩;⟨SMALL⟩;IGNORE % GREEK SMALL LETTER ETA WITH OXIA AND YPOGEGRAMMENI

⟨U1F74⟩ ⟨eta⟩;⟨GRAVE⟩;⟨SMALL⟩;IGNORE % GREEK SMALL LETTER ETA WITH VARIA

⟨U1FCA⟩ ⟨eta⟩;⟨GRAVE⟩;⟨CAPITAL⟩;IGNORE % GREEK CAPITAL LETTER ETA WITH VARIA

⟨U1FC2⟩ ⟨eta⟩;⟨VARIA+YPOGEGRAMMENI⟩;⟨SMALL⟩;IGNORE % GREEK SMALL LETTER ETA WITH VARIA AND YPOGEGRAMMENI

⟨U1FC6⟩ ⟨eta⟩;⟨PERISPOMENI⟩;⟨SMALL⟩;IGNORE % GREEK SMALL LETTER ETA WITH PERISPOMENI

⟨U1FC7⟩ ⟨eta⟩;⟨PERISPOMENI+YPOGEGRAMMENI⟩;⟨SMALL⟩;IGNORE % GREEK SMALL LETTER ETA WITH PERISPOMENI AND YPOGEGRAMMENI

⟨U03AE⟩ ⟨eta⟩;⟨TONOS⟩;⟨SMALL⟩;IGNORE % GREEK SMALL LETTER ETA WITH TONOS

⟨U0389⟩ ⟨eta⟩;⟨TONOS⟩;⟨CAPITAL⟩;IGNORE % GREEK CAPITAL LETTER ETA WITH TONOS

⟨U1FCC⟩ ⟨eta⟩;⟨PROSGEGRAMMENI⟩;⟨CAPITAL⟩;IGNORE % GREEK CAPITAL LETTER ETA WITH PROSGEGRAMMENI

⟨U1FC3⟩ ⟨eta⟩;⟨YPOGEGRAMMENI⟩;⟨SMALL⟩;IGNORE % GREEK SMALL LETTER ETA WITH YPOGEGRAMMENI

⟨U03B8⟩ ⟨theta⟩;⟨BLANK⟩;⟨SMALL⟩;IGNORE % GREEK SMALL LETTER THETA

⟨U0398⟩ ⟨theta⟩;⟨BLANK⟩;⟨CAPITAL⟩;IGNORE % GREEK CAPITAL LETTER THETA

⟨U03B9⟩ ⟨iota⟩;⟨BLANK⟩;⟨SMALL⟩;IGNORE % GREEK SMALL LETTER IOTA

⟨U0399⟩ ⟨iota⟩;⟨BLANK⟩;⟨CAPITAL⟩;IGNORE % GREEK CAPITAL LETTER IOTA

⟨U1F30⟩ ⟨iota⟩;⟨PSILI⟩;⟨SMALL⟩;IGNORE % GREEK SMALL LETTER IOTA WITH PSILI

⟨U1F38⟩ ⟨iota⟩;⟨PSILI⟩;⟨CAPITAL⟩;IGNORE % GREEK CAPITAL LETTER IOTA WITH PSILI

⟨U1F34⟩ ⟨iota⟩;⟨PSILI+OXIA⟩;⟨SMALL⟩;IGNORE % GREEK SMALL LETTER IOTA WITH PSILI AND OXIA

⟨U1F3C⟩ ⟨iota⟩;⟨PSILI+OXIA⟩;⟨CAPITAL⟩;IGNORE % GREEK CAPITAL LETTER IOTA WITH PSILI AND OXIA

⟨U1F32⟩ ⟨iota⟩;⟨PSILI+VARIA⟩;⟨SMALL⟩;IGNORE % GREEK SMALL LETTER IOTA WITH PSILI AND VARIA

⟨U1F3A⟩ ⟨iota⟩;⟨PSILI+VARIA⟩;⟨CAPITAL⟩;IGNORE % GREEK CAPITAL LETTER IOTA WITH PSILI AND VARIA

⟨U1F36⟩ ⟨iota⟩;⟨PSILI+PERISPOMENI⟩;⟨SMALL⟩;IGNORE % GREEK SMALL LETTER IOTA WITH PSILI AND PERISPOMENI

⟨U1F3E⟩ ⟨iota⟩;⟨PSILI+PERISPOMENI⟩;⟨CAPITAL⟩;IGNORE % GREEK CAPITAL LETTER IOTA WITH PSILI AND PERISPOMENI

⟨U1F31⟩ ⟨iota⟩;⟨DASIA⟩;⟨SMALL⟩;IGNORE % GREEK SMALL LETTER IOTA WITH DASIA

⟨U1F39⟩ ⟨iota⟩;⟨DASIA⟩;⟨CAPITAL⟩;IGNORE % GREEK CAPITAL LETTER IOTA WITH DASIA

⟨U1F35⟩ ⟨iota⟩;⟨DASIA+OXIA⟩;⟨SMALL⟩;IGNORE % GREEK SMALL LETTER IOTA WITH DASIA AND OXIA

⟨U1F3D⟩ ⟨iota⟩;⟨DASIA+OXIA⟩;⟨CAPITAL⟩;IGNORE % GREEK CAPITAL LETTER IOTA WITH DASIA AND OXIA
⟨U1F33⟩ ⟨iota⟩;⟨DASIA+VARIA⟩;⟨SMALL⟩;IGNORE % GREEK SMALL LETTER IOTA WITH DASIA AND VARIA
⟨U1F3B⟩ ⟨iota⟩;⟨DASIA+VARIA⟩;⟨CAPITAL⟩;IGNORE % GREEK CAPITAL LETTER IOTA WITH DASIA AND VARIA
⟨U1F37⟩ ⟨iota⟩;⟨DASIA+PERISPOMENI⟩;⟨SMALL⟩;IGNORE % GREEK SMALL LETTER IOTA WITH DASIA AND PERISPOMENI
⟨U1F3F⟩ ⟨iota⟩;⟨DASIA+PERISPOMENI⟩;⟨CAPITAL⟩;IGNORE % GREEK CAPITAL LETTER IOTA WITH DASIA AND PERISPOMENI
⟨U1F77⟩ ⟨iota⟩;⟨ACUTE⟩;⟨SMALL⟩;IGNORE % GREEK SMALL LETTER IOTA WITH OXIA
⟨U1FDB⟩ ⟨iota⟩;⟨ACUTE⟩;⟨CAPITAL⟩;IGNORE % GREEK CAPITAL LETTER IOTA WITH OXIA
⟨U1F76⟩ ⟨iota⟩;⟨GRAVE⟩;⟨SMALL⟩;IGNORE % GREEK SMALL LETTER IOTA WITH VARIA
⟨U1FDA⟩ ⟨iota⟩;⟨GRAVE⟩;⟨CAPITAL⟩;IGNORE % GREEK CAPITAL LETTER IOTA WITH VARIA
⟨U1FD6⟩ ⟨iota⟩;⟨PERISPOMENI⟩;⟨SMALL⟩;IGNORE % GREEK SMALL LETTER IOTA WITH PERISPOMENI
⟨U03AF⟩ ⟨iota⟩;⟨TONOS⟩;⟨SMALL⟩;IGNORE % GREEK SMALL LETTER IOTA WITH TONOS
⟨U038A⟩ ⟨iota⟩;⟨TONOS⟩;⟨CAPITAL⟩;IGNORE % GREEK CAPITAL LETTER IOTA WITH TONOS
⟨U1FD0⟩ ⟨iota⟩;⟨BREVE⟩;⟨SMALL⟩;IGNORE % GREEK SMALL LETTER IOTA WITH VRACHY
⟨U1FD8⟩ ⟨iota⟩;⟨BREVE⟩;⟨CAPITAL⟩;IGNORE % GREEK CAPITAL LETTER IOTA WITH VRACHY
⟨U03CA⟩ ⟨iota⟩;⟨DIAERESIS⟩;⟨SMALL⟩;IGNORE % GREEK SMALL LETTER IOTA WITH DIALYTIKA
⟨U03AA⟩ ⟨iota⟩;⟨DIAERESIS⟩;⟨CAPITAL⟩;IGNORE % GREEK CAPITAL LETTER IOTA WITH DIALYTIKA
⟨U1FD2⟩ ⟨iota⟩;⟨DIAERESIS+VARIA⟩;⟨SMALL⟩;IGNORE % GREEK SMALL LETTER IOTA WITH DIALYTIKA AND VARIA
⟨U1FD3⟩ ⟨iota⟩;⟨DIAERESIS+OXIA⟩;⟨SMALL⟩;IGNORE % GREEK SMALL LETTER IOTA WITH DIALYTIKA AND OXIA
⟨U1FD7⟩ ⟨iota⟩;⟨DIAERESIS+PERISPOMENI⟩;⟨SMALL⟩;IGNORE % GREEK SMALL LETTER IOTA WITH DIALYTIKA AND PERISPOMENI
⟨U0390⟩ ⟨iota⟩;⟨DIAERESIS+TONOS⟩;⟨SMALL⟩;IGNORE % GREEK SMALL LETTER IOTA WITH DIALYTIKA AND TONOS
⟨U1FD1⟩ ⟨iota⟩;⟨MACRON⟩;⟨SMALL⟩;IGNORE % GREEK SMALL LETTER IOTA WITH MACRON
⟨U1FD9⟩ ⟨iota⟩;⟨MACRON⟩;⟨CAPITAL⟩;IGNORE % GREEK CAPITAL LETTER IOTA WITH MACRON
⟨U03BA⟩ ⟨kappa⟩;⟨BLANK⟩;⟨SMALL⟩;IGNORE % GREEK SMALL LETTER KAPPA
⟨U039A⟩ ⟨kappa⟩;⟨BLANK⟩;⟨CAPITAL⟩;IGNORE % GREEK CAPITAL LETTER KAPPA
⟨U03BB⟩ ⟨lamda⟩;⟨BLANK⟩;⟨SMALL⟩;IGNORE % GREEK SMALL LETTER LAMDA
⟨U039B⟩ ⟨lamda⟩;⟨BLANK⟩;⟨CAPITAL⟩;IGNORE % GREEK CAPITAL LETTER LAMDA
⟨U03BC⟩ ⟨mu⟩;⟨BLANK⟩;⟨SMALL⟩;IGNORE % GREEK SMALL LETTER MU
⟨U039C⟩ ⟨mu⟩;⟨BLANK⟩;⟨CAPITAL⟩;IGNORE % GREEK CAPITAL LETTER MU
⟨U03BD⟩ ⟨nu⟩;⟨BLANK⟩;⟨SMALL⟩;IGNORE % GREEK SMALL LETTER NU
⟨U039D⟩ ⟨nu⟩;⟨BLANK⟩;⟨CAPITAL⟩;IGNORE % GREEK CAPITAL LETTER NU
⟨U03BE⟩ ⟨xi⟩;⟨BLANK⟩;⟨SMALL⟩;IGNORE % GREEK SMALL LETTER XI
⟨U039E⟩ ⟨xi⟩;⟨BLANK⟩;⟨CAPITAL⟩;IGNORE % GREEK CAPITAL LETTER XI
⟨U03BF⟩ ⟨omicron⟩;⟨BLANK⟩;⟨SMALL⟩;IGNORE % GREEK SMALL LETTER OMICRON
⟨U039F⟩ ⟨omicron⟩;⟨BLANK⟩;⟨CAPITAL⟩;IGNORE % GREEK CAPITAL LETTER OMICRON
⟨U1F40⟩ ⟨omicron⟩;⟨PSILI⟩;⟨SMALL⟩;IGNORE % GREEK SMALL LETTER OMICRON WITH PSILI
⟨U1F48⟩ ⟨omicron⟩;⟨PSILI⟩;⟨CAPITAL⟩;IGNORE % GREEK CAPITAL LETTER OMICRON WITH PSILI
⟨U1F44⟩ ⟨omicron⟩;⟨PSILI+OXIA⟩;⟨SMALL⟩;IGNORE % GREEK SMALL LETTER OMICRON WITH PSILI AND OXIA
⟨U1F4C⟩ ⟨omicron⟩;⟨PSILI+OXIA⟩;⟨CAPITAL⟩;IGNORE % GREEK CAPITAL LETTER OMICRON WITH PSILI AND OXIA
⟨U1F42⟩ ⟨omicron⟩;⟨PSILI+VARIA⟩;⟨SMALL⟩;IGNORE % GREEK SMALL LETTER OMICRON WITH PSILI AND VARIA
⟨U1F4A⟩ ⟨omicron⟩;⟨PSILI+VARIA⟩;⟨CAPITAL⟩;IGNORE % GREEK CAPITAL LETTER OMICRON WITH PSILI AND VARIA
⟨U1F41⟩ ⟨omicron⟩;⟨DASIA⟩;⟨SMALL⟩;IGNORE % GREEK SMALL LETTER OMICRON WITH DASIA
⟨U1F49⟩ ⟨omicron⟩;⟨DASIA⟩;⟨CAPITAL⟩;IGNORE % GREEK CAPITAL LETTER OMICRON WITH

DASIA
⟨U1F45⟩ ⟨omicron⟩;⟨DASIA+OXIA⟩;⟨SMALL⟩;IGNORE % GREEK SMALL LETTER OMICRON WITH DASIA AND OXIA
⟨U1F4D⟩ ⟨omicron⟩;⟨DASIA+OXIA⟩;⟨CAPITAL⟩;IGNORE % GREEK CAPITAL LETTER OMICRON WITH DASIA AND OXIA
⟨U1F43⟩ ⟨omicron⟩;⟨DASIA+VARIA⟩;⟨SMALL⟩;IGNORE % GREEK SMALL LETTER OMICRON WITH DASIA AND VARIA
⟨U1F4B⟩ ⟨omicron⟩;⟨DASIA+VARIA⟩;⟨CAPITAL⟩;IGNORE % GREEK CAPITAL LETTER OMICRON WITH DASIA AND VARIA
⟨U1F79⟩ ⟨omicron⟩;⟨ACUTE⟩;⟨SMALL⟩;IGNORE % GREEK SMALL LETTER OMICRON WITH OXIA
⟨U1FF9⟩ ⟨omicron⟩;⟨ACUTE⟩;⟨CAPITAL⟩;IGNORE % GREEK CAPITAL LETTER OMICRON WITH OXIA
⟨U1F78⟩ ⟨omicron⟩;⟨GRAVE⟩;⟨SMALL⟩;IGNORE % GREEK SMALL LETTER OMICRON WITH VARIA
⟨U1FF8⟩ ⟨omicron⟩;⟨GRAVE⟩;⟨CAPITAL⟩;IGNORE % GREEK CAPITAL LETTER OMICRON WITH VARIA
⟨U03CC⟩ ⟨omicron⟩;⟨TONOS⟩;⟨SMALL⟩;IGNORE % GREEK SMALL LETTER OMICRON WITH TONOS
⟨U038C⟩ ⟨omicron⟩;⟨TONOS⟩;⟨CAPITAL⟩;IGNORE % GREEK CAPITAL LETTER OMICRON WITH TONOS
⟨U03C0⟩ ⟨pi⟩;⟨BLANK⟩;⟨SMALL⟩;IGNORE % GREEK SMALL LETTER PI
⟨U03A0⟩ ⟨pi⟩;⟨BLANK⟩;⟨CAPITAL⟩;IGNORE % GREEK CAPITAL LETTER PI
⟨U03C1⟩ ⟨rho⟩;⟨BLANK⟩;⟨SMALL⟩;IGNORE % GREEK SMALL LETTER RHO
⟨U03A1⟩ ⟨rho⟩;⟨BLANK⟩;⟨CAPITAL⟩;IGNORE % GREEK CAPITAL LETTER RHO
⟨U1FE4⟩ ⟨rho⟩;⟨PSILI⟩;⟨SMALL⟩;IGNORE % GREEK SMALL LETTER RHO WITH PSILI
⟨U1FE5⟩ ⟨rho⟩;⟨DASIA⟩;⟨SMALL⟩;IGNORE % GREEK SMALL LETTER RHO WITH DASIA
⟨U1FEC⟩ ⟨rho⟩;⟨DASIA⟩;⟨CAPITAL⟩;IGNORE % GREEK CAPITAL LETTER RHO WITH DASIA
⟨U03C3⟩ ⟨sigma⟩;⟨BLANK⟩;⟨SMALL⟩;IGNORE % GREEK SMALL LETTER SIGMA
⟨U03A3⟩ ⟨sigma⟩;⟨BLANK⟩;⟨CAPITAL⟩;IGNORE % GREEK CAPITAL LETTER SIGMA
⟨U03C2⟩ ⟨finalsigma⟩;⟨BLANK⟩;⟨SMALL⟩;IGNORE % GREEK SMALL LETTER FINAL SIGMA
⟨U03C4⟩ ⟨tau⟩;⟨BLANK⟩;⟨SMALL⟩;IGNORE % GREEK SMALL LETTER TAU
⟨U03A4⟩ ⟨tau⟩;⟨BLANK⟩;⟨CAPITAL⟩;IGNORE % GREEK CAPITAL LETTER TAU
⟨U03C5⟩ ⟨upsilon⟩;⟨BLANK⟩;⟨SMALL⟩;IGNORE % GREEK SMALL LETTER UPSILON
⟨U03A5⟩ ⟨upsilon⟩;⟨BLANK⟩;⟨CAPITAL⟩;IGNORE % GREEK CAPITAL LETTER UPSILON
⟨U1F50⟩ ⟨upsilon⟩;⟨PSILI⟩;⟨SMALL⟩;IGNORE % GREEK SMALL LETTER UPSILON WITH PSILI
⟨U1F54⟩ ⟨upsilon⟩;⟨PSILI+OXIA⟩;⟨SMALL⟩;IGNORE % GREEK SMALL LETTER UPSILON WITH PSILI AND OXIA
⟨U1F52⟩ ⟨upsilon⟩;⟨PSILI+VARIA⟩;⟨SMALL⟩;IGNORE % GREEK SMALL LETTER UPSILON WITH PSILI AND VARIA
⟨U1F56⟩ ⟨upsilon⟩;⟨PSILI+PERISPOMENI⟩;⟨SMALL⟩;IGNORE % GREEK SMALL LETTER UPSILON WITH PSILI AND PERISPOMENI
⟨U1F51⟩ ⟨upsilon⟩;⟨DASIA⟩;⟨SMALL⟩;IGNORE % GREEK SMALL LETTER UPSILON WITH DASIA
⟨U1F59⟩ ⟨upsilon⟩;⟨DASIA⟩;⟨CAPITAL⟩;IGNORE % GREEK CAPITAL LETTER UPSILON WITH DASIA
⟨U1F55⟩ ⟨upsilon⟩;⟨DASIA+OXIA⟩;⟨SMALL⟩;IGNORE % GREEK SMALL LETTER UPSILON WITH DASIA AND OXIA
⟨U1F5D⟩ ⟨upsilon⟩;⟨DASIA+OXIA⟩;⟨CAPITAL⟩;IGNORE % GREEK CAPITAL LETTER UPSILON WITH DASIA AND OXIA
⟨U1F53⟩ ⟨upsilon⟩;⟨DASIA+VARIA⟩;⟨SMALL⟩;IGNORE % GREEK SMALL LETTER UPSILON WITH DASIA AND VARIA
⟨U1F5B⟩ ⟨upsilon⟩;⟨DASIA+VARIA⟩;⟨CAPITAL⟩;IGNORE % GREEK CAPITAL LETTER UPSILON WITH DASIA AND VARIA
⟨U1F57⟩ ⟨upsilon⟩;⟨DASIA+PERISPOMENI⟩;⟨SMALL⟩;IGNORE % GREEK SMALL LETTER UPSILON WITH DASIA AND PERISPOMENI
⟨U1F5F⟩ ⟨upsilon⟩;⟨DASIA+PERISPOMENI⟩;⟨CAPITAL⟩;IGNORE % GREEK CAPITAL LETTER UPSILON WITH DASIA AND PERISPOMENI
⟨U1F7B⟩ ⟨upsilon⟩;⟨ACUTE⟩;⟨SMALL⟩;IGNORE % GREEK SMALL LETTER UPSILON WITH OXIA
⟨U1FEB⟩ ⟨upsilon⟩;⟨ACUTE⟩;⟨CAPITAL⟩;IGNORE % GREEK CAPITAL LETTER UPSILON WITH OXIA
⟨U1F7A⟩ ⟨upsilon⟩;⟨GRAVE⟩;⟨SMALL⟩;IGNORE % GREEK SMALL LETTER UPSILON WITH VARIA
⟨U1FEA⟩ ⟨upsilon⟩;⟨GRAVE⟩;⟨CAPITAL⟩;IGNORE % GREEK CAPITAL LETTER UPSILON WITH VARIA

⟨U1FE6⟩ ⟨upsilon⟩;⟨PERISPOMENI⟩;⟨SMALL⟩;IGNORE % GREEK SMALL LETTER UPSILON WITH PERISPOMENI
⟨U03CD⟩ ⟨upsilon⟩;⟨TONOS⟩;⟨SMALL⟩;IGNORE % GREEK SMALL LETTER UPSILON WITH TONOS
⟨U038E⟩ ⟨upsilon⟩;⟨TONOS⟩;⟨CAPITAL⟩;IGNORE % GREEK CAPITAL LETTER UPSILON WITH TONOS
⟨U1FE0⟩ ⟨upsilon⟩;⟨BREVE⟩;⟨SMALL⟩;IGNORE % GREEK SMALL LETTER UPSILON WITH VRACHY
⟨U1FE8⟩ ⟨upsilon⟩;⟨BREVE⟩;⟨CAPITAL⟩;IGNORE % GREEK CAPITAL LETTER UPSILON WITH VRACHY
⟨U03CB⟩ ⟨upsilon⟩;⟨DIAERESIS⟩;⟨SMALL⟩;IGNORE % GREEK SMALL LETTER UPSILON WITH DIALYTIKA
⟨U03AB⟩ ⟨upsilon⟩;⟨DIAERESIS⟩;⟨CAPITAL⟩;IGNORE % GREEK CAPITAL LETTER UPSILON WITH DIALYTIKA
⟨U1FE2⟩ ⟨upsilon⟩;⟨DIAERESIS+VARIA⟩;⟨SMALL⟩;IGNORE % GREEK SMALL LETTER UPSILON WITH DIALYTIKA AND VARIA
⟨U1FE3⟩ ⟨upsilon⟩;⟨DIAERESIS+OXIA⟩;⟨SMALL⟩;IGNORE % GREEK SMALL LETTER UPSILON WITH DIALYTIKA AND OXIA
⟨U1FE7⟩ ⟨upsilon⟩;⟨DIAERESIS+PERISPOMENI⟩;⟨SMALL⟩;IGNORE % GREEK SMALL LETTER UPSILON WITH DIALYTIKA AND PERISPOMENI
⟨U03B0⟩ ⟨upsilon⟩;⟨DIAERESIS+TONOS⟩;⟨SMALL⟩;IGNORE % GREEK SMALL LETTER UPSILON WITH DIALYTIKA AND TONOS
⟨U1FE1⟩ ⟨upsilon⟩;⟨MACRON⟩;⟨SMALL⟩;IGNORE % GREEK SMALL LETTER UPSILON WITH MACRON
⟨U1FE9⟩ ⟨upsilon⟩;⟨MACRON⟩;⟨CAPITAL⟩;IGNORE % GREEK CAPITAL LETTER UPSILON WITH MACRON
⟨U03C6⟩ ⟨phi⟩;⟨BLANK⟩;⟨SMALL⟩;IGNORE % GREEK SMALL LETTER PHI
⟨U03A6⟩ ⟨phi⟩;⟨BLANK⟩;⟨CAPITAL⟩;IGNORE % GREEK CAPITAL LETTER PHI
⟨U03C7⟩ ⟨chi⟩;⟨BLANK⟩;⟨SMALL⟩;IGNORE % GREEK SMALL LETTER CHI
⟨U03A7⟩ ⟨chi⟩;⟨BLANK⟩;⟨CAPITAL⟩;IGNORE % GREEK CAPITAL LETTER CHI
⟨U03C8⟩ ⟨psi⟩;⟨BLANK⟩;⟨SMALL⟩;IGNORE % GREEK SMALL LETTER PSI
⟨U03A8⟩ ⟨psi⟩;⟨BLANK⟩;⟨CAPITAL⟩;IGNORE % GREEK CAPITAL LETTER PSI
⟨U03C9⟩ ⟨omega⟩;⟨BLANK⟩;⟨SMALL⟩;IGNORE % GREEK SMALL LETTER OMEGA
⟨U03A9⟩ ⟨omega⟩;⟨BLANK⟩;⟨CAPITAL⟩;IGNORE % GREEK CAPITAL LETTER OMEGA
⟨U1F60⟩ ⟨omega⟩;⟨PSILI⟩;⟨SMALL⟩;IGNORE % GREEK SMALL LETTER OMEGA WITH PSILI
⟨U1F68⟩ ⟨omega⟩;⟨PSILI⟩;⟨CAPITAL⟩;IGNORE % GREEK CAPITAL LETTER OMEGA WITH PSILI
⟨U1F64⟩ ⟨omega⟩;⟨PSILI+OXIA⟩;⟨SMALL⟩;IGNORE % GREEK SMALL LETTER OMEGA WITH PSILI AND OXIA
⟨U1F6C⟩ ⟨omega⟩;⟨PSILI+OXIA⟩;⟨CAPITAL⟩;IGNORE % GREEK CAPITAL LETTER OMEGA WITH PSILI AND OXIA
⟨U1FAC⟩ ⟨omega⟩;⟨PSILI+OXIA+PROSGEGRAMMENI⟩;⟨CAPITAL⟩;IGNORE % GREEK CAPITAL LETTER OMEGA WITH PSILI AND OXIA AND PROSGEGRAMMENI
⟨U1FA4⟩ ⟨omega⟩;⟨PSILI+OXIA+YPOGEGRAMMENI⟩;⟨SMALL⟩;IGNORE % GREEK SMALL LETTER OMEGA WITH PSILI AND OXIA AND YPOGEGRAMMENI
⟨U1F62⟩ ⟨omega⟩;⟨PSILI+VARIA⟩;⟨SMALL⟩;IGNORE % GREEK SMALL LETTER OMEGA WITH PSILI AND VARIA
⟨U1F6A⟩ ⟨omega⟩;⟨PSILI+VARIA⟩;⟨CAPITAL⟩;IGNORE % GREEK CAPITAL LETTER OMEGA WITH PSILI AND VARIA
⟨U1FAA⟩ ⟨omega⟩;⟨PSILI+VARIA+PROSGEGRAMMENI⟩;⟨CAPITAL⟩;IGNORE % GREEK CAPITAL LETTER OMEGA WITH PSILI AND VARIA AND PROSGEGRAMMENI
⟨U1FA2⟩ ⟨omega⟩;⟨PSILI+VARIA+YPOGEGRAMMENI⟩;⟨SMALL⟩;IGNORE % GREEK SMALL LETTER OMEGA WITH PSILI AND VARIA AND YPOGEGRAMMENI
⟨U1FAE⟩ ⟨omega⟩;⟨PSILI+PERISPOMENI+PROSGEGRAMMENI⟩;⟨CAPITAL⟩;IGNORE % GREEK CAPITAL LETTER OMEGA WITH PSILI AND PERISPOMENI AND PROSGEGRAMMENI
⟨U1FA6⟩ ⟨omega⟩;⟨PSILI+PERISPOMENI+YPOGEGRAMMENI⟩;⟨SMALL⟩;IGNORE % GREEK SMALL LETTER OMEGA WITH PSILI AND PERISPOMENI AND YPOGEGRAMMENI
⟨U1F66⟩ ⟨omega⟩;⟨PSILI+PERISPOMENI⟩;⟨SMALL⟩;IGNORE % GREEK SMALL LETTER OMEGA WITH PSILI AND PERISPOMENI
⟨U1F6E⟩ ⟨omega⟩;⟨PSILI+PERISPOMENI⟩;⟨CAPITAL⟩;IGNORE % GREEK CAPITAL LETTER OMEGA WITH PSILI AND PERISPOMENI
⟨U1FA8⟩ ⟨omega⟩;⟨PSILI+PROSGEGRAMMENI⟩;⟨CAPITAL⟩;IGNORE % GREEK CAPITAL LETTER

OMEGA WITH PSILI AND PROSGEGRAMMENI
⟨U1FA0⟩ ⟨omega⟩;⟨PSILI+YPOGEGRAMMENI⟩;⟨SMALL⟩;IGNORE % GREEK SMALL LETTER OMEGA WITH PSILI AND YPOGEGRAMMENI
⟨U1F61⟩ ⟨omega⟩;⟨DASIA⟩;⟨SMALL⟩;IGNORE % GREEK SMALL LETTER OMEGA WITH DASIA
⟨U1F69⟩ ⟨omega⟩;⟨DASIA⟩;⟨CAPITAL⟩;IGNORE % GREEK CAPITAL LETTER OMEGA WITH DASIA
⟨U1F65⟩ ⟨omega⟩;⟨DASIA+OXIA⟩;⟨SMALL⟩;IGNORE % GREEK SMALL LETTER OMEGA WITH DASIA AND OXIA
⟨U1F6D⟩ ⟨omega⟩;⟨DASIA+OXIA⟩;⟨CAPITAL⟩;IGNORE % GREEK CAPITAL LETTER OMEGA WITH DASIA AND OXIA
⟨U1FAD⟩ ⟨omega⟩;⟨DASIA+OXIA+PROSGEGRAMMENI⟩;⟨CAPITAL⟩;IGNORE % GREEK CAPITAL LETTER OMEGA WITH DASIA AND OXIA AND PROSGEGRAMMENI
⟨U1FA5⟩ ⟨omega⟩;⟨DASIA+OXIA+YPOGEGRAMMENI⟩;⟨SMALL⟩;IGNORE % GREEK SMALL LETTER OMEGA WITH DASIA AND OXIA AND YPOGEGRAMMENI
⟨U1F63⟩ ⟨omega⟩;⟨DASIA+VARIA⟩;⟨SMALL⟩;IGNORE % GREEK SMALL LETTER OMEGA WITH DASIA AND VARIA
⟨U1F6B⟩ ⟨omega⟩;⟨DASIA+VARIA⟩;⟨CAPITAL⟩;IGNORE % GREEK CAPITAL LETTER OMEGA WITH DASIA AND VARIA
⟨U1FAB⟩ ⟨omega⟩;⟨DASIA+VARIA+PROSGEGRAMMENI⟩;⟨CAPITAL⟩;IGNORE % GREEK CAPITAL LETTER OMEGA WITH DASIA AND VARIA AND PROSGEGRAMMENI
⟨U1FA3⟩ ⟨omega⟩;⟨DASIA+VARIA+YPOGEGRAMMENI⟩;⟨SMALL⟩;IGNORE % GREEK SMALL LETTER OMEGA WITH DASIA AND VARIA AND YPOGEGRAMMENI
⟨U1F67⟩ ⟨omega⟩;⟨DASIA+PERISPOMENI⟩;⟨SMALL⟩;IGNORE % GREEK SMALL LETTER OMEGA WITH DASIA AND PERISPOMENI
⟨U1F6F⟩ ⟨omega⟩;⟨DASIA+PERISPOMENI⟩;⟨CAPITAL⟩;IGNORE % GREEK CAPITAL LETTER OMEGA WITH DASIA AND PERISPOMENI
⟨U1FAF⟩ ⟨omega⟩;⟨DASIA+PERISPOMENI+PROSGEGRAMMENI⟩;⟨CAPITAL⟩;IGNORE % GREEK CAPITAL LETTER OMEGA WITH DASIA AND PERISPOMENI AND PROSGEGRAMMENI
⟨U1FA7⟩ ⟨omega⟩;⟨DASIA+PERISPOMENI+YPOGEGRAMMENI⟩;⟨SMALL⟩;IGNORE % GREEK SMALL LETTER OMEGA WITH DASIA AND PERISPOMENI AND YPOGEGRAMMENI
⟨U1FA9⟩ ⟨omega⟩;⟨DASIA+PROSGEGRAMMENI⟩;⟨CAPITAL⟩;IGNORE % GREEK CAPITAL LETTER OMEGA WITH DASIA AND PROSGEGRAMMENI
⟨U1FA1⟩ ⟨omega⟩;⟨DASIA+YPOGEGRAMMENI⟩;⟨SMALL⟩;IGNORE % GREEK SMALL LETTER OMEGA WITH DASIA AND YPOGEGRAMMENI
⟨U1F7D⟩ ⟨omega⟩;⟨ACUTE⟩;⟨SMALL⟩;IGNORE % GREEK SMALL LETTER OMEGA WITH OXIA
⟨U1FFB⟩ ⟨omega⟩;⟨ACUTE⟩;⟨CAPITAL⟩;IGNORE % GREEK CAPITAL LETTER OMEGA WITH OXIA
⟨U1FF4⟩ ⟨omega⟩;⟨OXIA+YPOGEGRAMMENI⟩;⟨SMALL⟩;IGNORE % GREEK SMALL LETTER OMEGA WITH OXIA AND YPOGEGRAMMENI
⟨U1F7C⟩ ⟨omega⟩;⟨GRAVE⟩;⟨SMALL⟩;IGNORE % GREEK SMALL LETTER OMEGA WITH VARIA
⟨U1FFA⟩ ⟨omega⟩;⟨GRAVE⟩;⟨CAPITAL⟩;IGNORE % GREEK CAPITAL LETTER OMEGA WITH VARIA
⟨U1FF2⟩ ⟨omega⟩;⟨VARIA+YPOGEGRAMMENI⟩;⟨SMALL⟩;IGNORE % GREEK SMALL LETTER OMEGA WITH VARIA AND YPOGEGRAMMENI
⟨U1FF6⟩ ⟨omega⟩;⟨PERISPOMENI⟩;⟨SMALL⟩;IGNORE % GREEK SMALL LETTER OMEGA WITH PERISPOMENI
⟨U1FF7⟩ ⟨omega⟩;⟨PERISPOMENI+YPOGEGRAMMENI⟩;⟨SMALL⟩;IGNORE % GREEK SMALL LETTER OMEGA WITH PERISPOMENI AND YPOGEGRAMMENI
⟨U03CE⟩ ⟨omega⟩;⟨TONOS⟩;⟨SMALL⟩;IGNORE % GREEK SMALL LETTER OMEGA WITH TONOS
⟨U038F⟩ ⟨omega⟩;⟨TONOS⟩;⟨CAPITAL⟩;IGNORE % GREEK CAPITAL LETTER OMEGA WITH TONOS
⟨U1FFC⟩ ⟨omega⟩;⟨PROSGEGRAMMENI⟩;⟨CAPITAL⟩;IGNORE % GREEK CAPITAL LETTER OMEGA WITH PROSGEGRAMMENI
⟨U1FF3⟩ ⟨omega⟩;⟨YPOGEGRAMMENI⟩;⟨SMALL⟩;IGNORE % GREEK SMALL LETTER OMEGA WITH YPOGEGRAMMENI

section EORCyrillic
⟨U0430⟩ ⟨acyril⟩;⟨BLANK⟩;⟨SMALL⟩;IGNORE % CYRILLIC SMALL LETTER A
⟨U0410⟩ ⟨acyril⟩;⟨BLANK⟩;⟨CAPITAL⟩;IGNORE % CYRILLIC CAPITAL LETTER A
⟨U0431⟩ ⟨be⟩;⟨BLANK⟩;⟨SMALL⟩;IGNORE % CYRILLIC SMALL LETTER BE
⟨U0411⟩ ⟨be⟩;⟨BLANK⟩;⟨CAPITAL⟩;IGNORE % CYRILLIC CAPITAL LETTER BE
⟨U0432⟩ ⟨ve⟩;⟨BLANK⟩;⟨SMALL⟩;IGNORE % CYRILLIC SMALL LETTER VE

⟨U0412⟩ ⟨ve⟩;⟨BLANK⟩;⟨CAPITAL⟩;IGNORE % CYRILLIC CAPITAL LETTER VE
⟨U0433⟩ ⟨ge⟩;⟨BLANK⟩;⟨SMALL⟩;IGNORE % CYRILLIC SMALL LETTER GHE
⟨U0413⟩ ⟨ge⟩;⟨BLANK⟩;⟨CAPITAL⟩;IGNORE % CYRILLIC CAPITAL LETTER GHE
⟨U0453⟩ ⟨gje⟩;⟨BLANK⟩;⟨SMALL⟩;IGNORE % CYRILLIC SMALL LETTER GJE
⟨U0403⟩ ⟨gje⟩;⟨BLANK⟩;⟨CAPITAL⟩;IGNORE % CYRILLIC CAPITAL LETTER GJE
⟨U0491⟩ ⟨geupturn⟩;⟨BLANK⟩;⟨SMALL⟩;IGNORE % CYRILLIC SMALL LETTER GHE WITH UPTURN
⟨U0490⟩ ⟨geupturn⟩;⟨BLANK⟩;⟨CAPITAL⟩;IGNORE % CYRILLIC CAPITAL LETTER GHE WITH UPTURN
⟨U0434⟩ ⟨de⟩;⟨BLANK⟩;⟨SMALL⟩;IGNORE % CYRILLIC SMALL LETTER DE
⟨U0414⟩ ⟨de⟩;⟨BLANK⟩;⟨CAPITAL⟩;IGNORE % CYRILLIC CAPITAL LETTER DE
⟨U0452⟩ ⟨dje⟩;⟨BLANK⟩;⟨SMALL⟩;IGNORE % CYRILLIC SMALL LETTER DJE
⟨U0402⟩ ⟨dje⟩;⟨BLANK⟩;⟨CAPITAL⟩;IGNORE % CYRILLIC CAPITAL LETTER DJE
⟨U0435⟩ ⟨ie⟩;⟨BLANK⟩;⟨SMALL⟩;IGNORE % CYRILLIC SMALL LETTER IE
⟨U0415⟩ ⟨ie⟩;⟨BLANK⟩;⟨CAPITAL⟩;IGNORE % CYRILLIC CAPITAL LETTER IE
⟨U0451⟩ ⟨io⟩;⟨BLANK⟩;⟨SMALL⟩;IGNORE % CYRILLIC SMALL LETTER IO
⟨U0401⟩ ⟨io⟩;⟨BLANK⟩;⟨CAPITAL⟩;IGNORE % CYRILLIC CAPITAL LETTER IO
⟨U0454⟩ ⟨ecyril⟩;⟨BLANK⟩;⟨SMALL⟩;IGNORE % CYRILLIC SMALL LETTER UKRAINIAN IE
⟨U0404⟩ ⟨ecyril⟩;⟨BLANK⟩;⟨CAPITAL⟩;IGNORE % CYRILLIC CAPITAL LETTER UKRAINIAN IE
⟨U0436⟩ ⟨zhe⟩;⟨BLANK⟩;⟨SMALL⟩;IGNORE % CYRILLIC SMALL LETTER ZHE
⟨U0416⟩ ⟨zhe⟩;⟨BLANK⟩;⟨CAPITAL⟩;IGNORE % CYRILLIC CAPITAL LETTER ZHE
⟨U0437⟩ ⟨ze⟩;⟨BLANK⟩;⟨SMALL⟩;IGNORE % CYRILLIC SMALL LETTER ZE
⟨U0417⟩ ⟨ze⟩;⟨BLANK⟩;⟨CAPITAL⟩;IGNORE % CYRILLIC CAPITAL LETTER ZE
⟨U0455⟩ ⟨dze⟩;⟨BLANK⟩;⟨SMALL⟩;IGNORE % CYRILLIC SMALL LETTER DZE
⟨U0405⟩ ⟨dze⟩;⟨BLANK⟩;⟨CAPITAL⟩;IGNORE % CYRILLIC CAPITAL LETTER DZE
⟨U0438⟩ ⟨ii⟩;⟨BLANK⟩;⟨SMALL⟩;IGNORE % CYRILLIC SMALL LETTER I
⟨U0418⟩ ⟨ii⟩;⟨BLANK⟩;⟨CAPITAL⟩;IGNORE % CYRILLIC CAPITAL LETTER I
⟨U0456⟩ ⟨icyril⟩;⟨BLANK⟩;⟨SMALL⟩;IGNORE % CYRILLIC SMALL LETTER BYELORUSSIAN-UKRAINIAN I
⟨U0406⟩ ⟨icyril⟩;⟨BLANK⟩;⟨CAPITAL⟩;IGNORE % CYRILLIC CAPITAL LETTER BYELORUSSIAN-UKRAINIAN I
⟨U0457⟩ ⟨yi⟩;⟨BLANK⟩;⟨SMALL⟩;IGNORE % CYRILLIC SMALL LETTER YI
⟨U0407⟩ ⟨yi⟩;⟨BLANK⟩;⟨CAPITAL⟩;IGNORE % CYRILLIC CAPITAL LETTER YI
⟨U0439⟩ ⟨iibreve⟩;⟨BLANK⟩;⟨SMALL⟩;IGNORE % CYRILLIC SMALL LETTER SHORT I
⟨U0419⟩ ⟨iibreve⟩;⟨BLANK⟩;⟨CAPITAL⟩;IGNORE % CYRILLIC CAPITAL LETTER SHORT I
⟨U0458⟩ ⟨je⟩;⟨BLANK⟩;⟨SMALL⟩;IGNORE % CYRILLIC SMALL LETTER JE
⟨U0408⟩ ⟨je⟩;⟨BLANK⟩;⟨CAPITAL⟩;IGNORE % CYRILLIC CAPITAL LETTER JE
⟨U043A⟩ ⟨ka⟩;⟨BLANK⟩;⟨SMALL⟩;IGNORE % CYRILLIC SMALL LETTER KA
⟨U041A⟩ ⟨ka⟩;⟨BLANK⟩;⟨CAPITAL⟩;IGNORE % CYRILLIC CAPITAL LETTER KA
⟨U045C⟩ ⟨kje⟩;⟨BLANK⟩;⟨SMALL⟩;IGNORE % CYRILLIC SMALL LETTER KJE
⟨U040C⟩ ⟨kje⟩;⟨BLANK⟩;⟨CAPITAL⟩;IGNORE % CYRILLIC CAPITAL LETTER KJE
⟨U043B⟩ ⟨el⟩;⟨BLANK⟩;⟨SMALL⟩;IGNORE % CYRILLIC SMALL LETTER EL
⟨U041B⟩ ⟨el⟩;⟨BLANK⟩;⟨CAPITAL⟩;IGNORE % CYRILLIC CAPITAL LETTER EL
⟨U0459⟩ ⟨lje⟩;⟨BLANK⟩;⟨SMALL⟩;IGNORE % CYRILLIC SMALL LETTER LJE
⟨U0409⟩ ⟨lje⟩;⟨BLANK⟩;⟨CAPITAL⟩;IGNORE % CYRILLIC CAPITAL LETTER LJE
⟨U043C⟩ ⟨em⟩;⟨BLANK⟩;⟨SMALL⟩;IGNORE % CYRILLIC SMALL LETTER EM
⟨U041C⟩ ⟨em⟩;⟨BLANK⟩;⟨CAPITAL⟩;IGNORE % CYRILLIC CAPITAL LETTER EM
⟨U043D⟩ ⟨en⟩;⟨BLANK⟩;⟨SMALL⟩;IGNORE % CYRILLIC SMALL LETTER EN
⟨U041D⟩ ⟨en⟩;⟨BLANK⟩;⟨CAPITAL⟩;IGNORE % CYRILLIC CAPITAL LETTER EN
⟨U045A⟩ ⟨nje⟩;⟨BLANK⟩;⟨SMALL⟩;IGNORE % CYRILLIC SMALL LETTER NJE
⟨U040A⟩ ⟨nje⟩;⟨BLANK⟩;⟨CAPITAL⟩;IGNORE % CYRILLIC CAPITAL LETTER NJE
⟨U043E⟩ ⟨ocyril⟩;⟨BLANK⟩;⟨SMALL⟩;IGNORE % CYRILLIC SMALL LETTER O
⟨U041E⟩ ⟨ocyril⟩;⟨BLANK⟩;⟨CAPITAL⟩;IGNORE % CYRILLIC CAPITAL LETTER O
⟨U043F⟩ ⟨pecyril⟩;⟨BLANK⟩;⟨SMALL⟩;IGNORE % CYRILLIC SMALL LETTER PE
⟨U041F⟩ ⟨pecyril⟩;⟨BLANK⟩;⟨CAPITAL⟩;IGNORE % CYRILLIC CAPITAL LETTER PE
⟨U0440⟩ ⟨er⟩;⟨BLANK⟩;⟨SMALL⟩;IGNORE % CYRILLIC SMALL LETTER ER
⟨U0420⟩ ⟨er⟩;⟨BLANK⟩;⟨CAPITAL⟩;IGNORE % CYRILLIC CAPITAL LETTER ER
⟨U0441⟩ ⟨es⟩;⟨BLANK⟩;⟨SMALL⟩;IGNORE % CYRILLIC SMALL LETTER ES
⟨U0421⟩ ⟨es⟩;⟨BLANK⟩;⟨CAPITAL⟩;IGNORE % CYRILLIC CAPITAL LETTER ES

```
⟨U0442⟩ ⟨te⟩;⟨BLANK⟩;⟨SMALL⟩;IGNORE % CYRILLIC SMALL LETTER TE
⟨U0422⟩ ⟨te⟩;⟨BLANK⟩;⟨CAPITAL⟩;IGNORE % CYRILLIC CAPITAL LETTER TE
⟨U045B⟩ ⟨tshe⟩;⟨BLANK⟩;⟨SMALL⟩;IGNORE % CYRILLIC SMALL LETTER TSHE
⟨U040B⟩ ⟨tshe⟩;⟨BLANK⟩;⟨CAPITAL⟩;IGNORE % CYRILLIC CAPITAL LETTER TSHE
⟨U0443⟩ ⟨ucyril⟩;⟨BLANK⟩;⟨SMALL⟩;IGNORE % CYRILLIC SMALL LETTER U
⟨U0423⟩ ⟨ucyril⟩;⟨BLANK⟩;⟨CAPITAL⟩;IGNORE % CYRILLIC CAPITAL LETTER U
⟨U045E⟩ ⟨ucyrilbreve⟩;⟨BLANK⟩;⟨SMALL⟩;IGNORE % CYRILLIC SMALL LETTER SHORT U
⟨U040E⟩ ⟨ucyrilbreve⟩;⟨BLANK⟩;⟨CAPITAL⟩;IGNORE % CYRILLIC CAPITAL LETTER SHORT U
⟨U0444⟩ ⟨ef⟩;⟨BLANK⟩;⟨SMALL⟩;IGNORE % CYRILLIC SMALL LETTER EF
⟨U0424⟩ ⟨ef⟩;⟨BLANK⟩;⟨CAPITAL⟩;IGNORE % CYRILLIC CAPITAL LETTER EF
⟨U0445⟩ ⟨kha⟩;⟨BLANK⟩;⟨SMALL⟩;IGNORE % CYRILLIC SMALL LETTER HA
⟨U0425⟩ ⟨kha⟩;⟨BLANK⟩;⟨CAPITAL⟩;IGNORE % CYRILLIC CAPITAL LETTER HA
⟨U0446⟩ ⟨tse⟩;⟨BLANK⟩;⟨SMALL⟩;IGNORE % CYRILLIC SMALL LETTER TSE
⟨U0426⟩ ⟨tse⟩;⟨BLANK⟩;⟨CAPITAL⟩;IGNORE % CYRILLIC CAPITAL LETTER TSE
⟨U0447⟩ ⟨che⟩;⟨BLANK⟩;⟨SMALL⟩;IGNORE % CYRILLIC SMALL LETTER CHE
⟨U0427⟩ ⟨che⟩;⟨BLANK⟩;⟨CAPITAL⟩;IGNORE % CYRILLIC CAPITAL LETTER CHE
⟨U045F⟩ ⟨dzhe⟩;⟨BLANK⟩;⟨SMALL⟩;IGNORE % CYRILLIC SMALL LETTER DZHE
⟨U040F⟩ ⟨dzhe⟩;⟨BLANK⟩;⟨CAPITAL⟩;IGNORE % CYRILLIC CAPITAL LETTER DZHE
⟨U0448⟩ ⟨sha⟩;⟨BLANK⟩;⟨SMALL⟩;IGNORE % CYRILLIC SMALL LETTER SHA
⟨U0428⟩ ⟨sha⟩;⟨BLANK⟩;⟨CAPITAL⟩;IGNORE % CYRILLIC CAPITAL LETTER SHA
⟨U0449⟩ ⟨shcha⟩;⟨BLANK⟩;⟨SMALL⟩;IGNORE % CYRILLIC SMALL LETTER SHCHA
⟨U0429⟩ ⟨shcha⟩;⟨BLANK⟩;⟨CAPITAL⟩;IGNORE % CYRILLIC CAPITAL LETTER SHCHA
⟨U044A⟩ ⟨hard⟩;⟨BLANK⟩;⟨SMALL⟩;IGNORE % CYRILLIC SMALL LETTER HARD SIGN
⟨U042A⟩ ⟨hard⟩;⟨BLANK⟩;⟨CAPITAL⟩;IGNORE % CYRILLIC CAPITAL LETTER HARD SIGN
⟨U044B⟩ ⟨yeri⟩;⟨BLANK⟩;⟨SMALL⟩;IGNORE % CYRILLIC SMALL LETTER YERU
⟨U042B⟩ ⟨yeri⟩;⟨BLANK⟩;⟨CAPITAL⟩;IGNORE % CYRILLIC CAPITAL LETTER YERU
⟨U044C⟩ ⟨soft⟩;⟨BLANK⟩;⟨SMALL⟩;IGNORE % CYRILLIC SMALL LETTER SOFT SIGN
⟨U042C⟩ ⟨soft⟩;⟨BLANK⟩;⟨CAPITAL⟩;IGNORE % CYRILLIC CAPITAL LETTER SOFT SIGN
⟨U044D⟩ ⟨ecyrilrev⟩;⟨BLANK⟩;⟨SMALL⟩;IGNORE % CYRILLIC SMALL LETTER E
⟨U042D⟩ ⟨ecyrilrev⟩;⟨BLANK⟩;⟨CAPITAL⟩;IGNORE % CYRILLIC CAPITAL LETTER E
⟨U044E⟩ ⟨iu⟩;⟨BLANK⟩;⟨SMALL⟩;IGNORE % CYRILLIC SMALL LETTER YU
⟨U042E⟩ ⟨iu⟩;⟨BLANK⟩;⟨CAPITAL⟩;IGNORE % CYRILLIC CAPITAL LETTER YU
⟨U044F⟩ ⟨ia⟩;⟨BLANK⟩;⟨SMALL⟩;IGNORE % CYRILLIC SMALL LETTER YA
⟨U042F⟩ ⟨ia⟩;⟨BLANK⟩;⟨CAPITAL⟩;IGNORE % CYRILLIC CAPITAL LETTER YA

%% Move the collating symbols to the start of the tailored table
reorder_section_after EORCollatingSymbols ⟨U0000⟩

%% Move this table after the space characters
reorder_section_after EORDigits ⟨U0020⟩

%% Latin after Digits
reorder_section_after EORLatin ⟨U2182⟩

%% Greek after Latin
reorder_section_after EORGreek ⟨U00FE⟩

%% Cyrillic after Greek
reorder_section_after EORCyrillic ⟨U1FF3⟩

%% Special characters are to be ordered in the sequence of ISO/IEC FCD 14651

order_start EORDigits; forward; forward; forward; forward, position
order_start EORLatin; forward; forward; forward; forward, position
order_start EORGreek; forward; forward; forward; forward, position
order_start EORCyrillic; forward; forward; forward; forward, position
order_end
```

Annex A (informative) Exposition of relevant principles

A.0 Introduction

This annex aims to present the information inherent in section 7 in a more accessible form for those who are interested in the principles guiding the composition of the table. Those readers not concerned with implementation details may take this more traditional treatment of the matter as an authoritative interpretatation of the information which is contained in the body of this European Standard.

A.1 Definitions

For the purpose of this annex, the following definitions apply in addition to those in the body of this European Standard.

A.1.1
character
unit of information used for the organization, control, or representation of textual data (ISO/IEC 10646–1)[6]

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

A.1.2
digit
one of the characters 0 1 2 3 4 5 6 7 8 9

A.1.3
letter
character used to represent (either alone or in combination) sounds or sequences of sounds of a natural language in writing.[7] Here equivalent to all characters of the *Multilingual European Subset No 2*[8] whose name contains one of the words LETTER or LIGATURE

A.1.4
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
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 ς τ T υ Y φ Φ χ X ψ Ψ ω Ω

---

[6] ⟨NOTE⟩ **Must still be checked**⟨/NOTE⟩

[7] ⟨NOTE⟩ This definition has been thus termed to exclude punctuation marks from being interpreted as letters. It is in line with standard linguistic terminology as exemplified, e. g., by Hadumod Bußmann: *Lexikon der Sprachwissenschaft*, Kröner: Stuttgart, [2]1990. Cf. also the relevant sections of the first draft of the ISSS Workshop on *Repertoires of characters used for writing the indigenous languages of Europe* ⟨/NOTE⟩

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

NOTE The final sigma ς is treated as a basic letter at the request of ΕΛΟΤ

a А б Б в В г Г г Г д Д е Е є Є ж Ж s S з З и И й Й i I j J 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 ь Ь э Э ю Ю я Я [9]

A.1.5
diacritical mark
any of a number of recurring graphical strutures placed over, under, next to, or through a basic letter which does not significantly modify the shape of the basic letter itself[10] and which in combination with that basic letter is a valid letter. These structures modify meaning or pronunciation or some other feature of the basic letter. The diacritical marks which are relevant to this European Standard are listed in Table 1

A.1.6
letter with diacritical marks
letter which is constructed as the combination between a basic letter and a diacritical mark

A.1.7
equivalent letter form
character created by joining two or more distinct basic letters or two or more letters with diacritical marks or any combination of these. The equivalent letter forms which are relevant to this European Standard are listed in Table 3[11]

A.1.8
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

A.1.9
capital letter
one of the basic letters

ABCDEFGHIJKLMNOPQRSTUVWXYZÞ

ΑΒΓΔΕΖΗΘΙΚΛΜΝΞΟΠΡΣΤΥΦΧΨΩ

АБВГГДЕЄЖSЗЗИЙІЈКЛЛЬМНЊОПРСТЋУФХЦЧШЩЪЫЬЭЮЯ

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

A.1.10
small letter
letter which is not a capital letter

A.1.11
special character
character that is neither a letter nor a digit

---

[9] ⟨NOTE⟩ This list is based on a list transmitted by John Clews. Relevant references include the Толковый словарь живаго великорускаго языка by Владимир Даль (С. Петербургъ , 1880) ⟨/NOTE⟩

[10] ⟨NOTE⟩ As Mr. van Wingen correctly pointed out, some diacritics as the diaeresis on ï do modify the basic letters⟨/NOTE⟩

[11] ⟨NOTE⟩ To cater for user expectations this European Standard consciously avoids the distinction between linguistic and non-linguistic ligatures⟨/NOTE⟩

A.1.12
character string
series of characters in sequence. Hereafter often referenced as string

A.1.13
ordering
act of bringing character strings into a well-defined sequence according to a string comparison specification (ISO FDIS 12199)

A.1.14
space character
one of the characters `SPACE` and `NO-BREAK SPACE`

A.2 Preparatory procedures

A.2.1 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.

A.2.2 Exposition

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 capital letters to small letters 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.

A.2.3 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.

A.3 The multilevel ordering procedure

A.3.1 General principles

This European Standard defines in this annex a multilevel ordering procedure whose results are identical to those produced by the application of the rules of the body of this standard.

Multilevel ordering procedure 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, finally the *fourth ordering level.' 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

A.3.2 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. Forms used primarily for emphasis are should likewise follow after more basic forms

A.3.3.3 Concept of canonical equivalence

The concept of canonical equivalence as defined by Unicode 2.0in section 3.6 is adopted throughout.

NOTE *Canonical equivalence* is not a concept of ISO/IEC 10646–1, but user expectations will demand it be observed as soon as combining characters are allowed for. While no such characters appear in the MES–2, they do so in MES–3. Some of the consequences of the application of this rule are somewhat unfortunate

A.3.3 Rules (valid throughout)

A.3.3.1 Ordering by script

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

A.3.3.2 Equivalent letter forms

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

A.4 Primary ordering level

A.4.1 Validity

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

A.4.2 Equivalent or ignored characters

A.4.2.1 Capital and small letters

Capital and small forms of the same letter are treated as equivalent

A.4.2.2 Modified letters

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

A.4.2.3 Letters with diacritical marks

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

A.4.2.4 Special characters

Special characters are ignored

A.4.3 Ordering sequences

A.4.3.1 Digits

Digits are to be ordered in the following sequence:
    0 1 2 3 4 5 6 7 8 9

A.4.3.2 Latin script

Basic Latin letters are to be ordered in the following sequence:
   a b c d e f g h i j k l mn o p q r st u v w x y z þ

A.4.3.3 Greek script

Basic Greek letters are to be ordered in the following sequence:
   α β γ δ ε ζ η θ ι κ λ μ ν ξ ο π ρ σ ς τ υ φ χ ψ ω

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

A.4.3.4 Cyrillic script

Basic Cyrillic letters are to be ordered in the following sequence:
   а б в г г̇ д е є ж з з и й і ј к л љ м н њ о п р с т ћ у ф х ц ч ш щ ъ ы ь э ю я

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

A.5 Secondary ordering level

A.5.1 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.

A.5.2 Equivalent or ignored characters

A.5.2.1 Capital and small letters

Capital and small forms of the same letter are treated as equivalent.

A.5.2.2 Special characters

Special characters are ignored.

A.5.3 Ordering sequences

A.5.3.1 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 specified by Table 4.

A.5.3.2 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.[12] In practice, this results for MES–2 in the sequence indicated in Table 2.

A.6 Tertiary ordering level

A.6.1 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 capital and small letters as equivalent.

A.6.2 Ignored characters

Special characters are ignored.

A.6.3 Ordering sequences

A.6.3.1 Capitalization

Small letters are ordered before the corresponding capital ones.[13]

A.7 Fourth ordering level

A.7.1 No unique sequence after the tertiary ordering level

If the tertiary ordering level likewise does not result in a unique sequence of strings, the fourth ordering level is invoked. It takes special characters into account.

A.7.2.Sequence of special characters

The special characters of the *Multilingual European Subset No 2* are ordered in the sequence of the default tailorable template of ISO/IEC FCD 14651, that is at presence in the sequence of their appearance in ISO/IEC 10646–1 and relevant appendices.

---

[12] ⟨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 å⟨/NOTE⟩

[13] ⟨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 small 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 small before capital, one (Denmark) capital before small and four (Great Britain, Finland, France and Norway) do not have a fixed rule ⟨/NOTE⟩

A.7.3 Equivalence

Two strings between which after the fourth 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 C: *Ordering by position and by style*

Specific ordering sequences

A.8.1 Diacritical marks

A.8.1.1 Diacritical marks

Table 1 – Diacritical marks[14]

Shape[15]	Corresponding diacritical mark in ISO/IEC 10646[16]	Alternative names[17]
	Diacritical marks for the Greek script	
'	U1FBF PSILI	spiritus lenis
'	U1FFE DASIA	spiritus asper
´	U1FFD OXIA[18]	
`	U1FEF VARIA	
˘	U0306 COMBINING BREVE	VRACHY
~	U0342 COMBINING GREEK PERISPOMENI	
'	U0384 TONOS	
ι	U1FBE PROSGEGRAMMENI	iota adscriptum[19]
	U0345 COMBINING GREEK YPOGEGRAMMENI	iota subscriptum[20]
¨	U0308 COMBINING DIAERESIS	DIALYTICA
‾	U0304 COMBINING MACRON	Greek macron, length
	Diacritical marks for Latin and Cyrillic scripts	
´	U0301 COMBINING ACUTE ACCENT	
`	U0300 COMBINING GRAVE ACCENT	
˘	U0306 COMBINING BREVE	
^	U0302 COMBINING CIRCUMFLEX ACCENT	
ˇ	U030C COMBINING CARON	
°	U030A COMBINING RING ABOVE	
¨	U0308 COMBINING DIAERESIS	umlaut, trema[21]
˝	U030B COMBINING DOUBLE ACUTE	
~	U0303 COMBINING TILDE	
·	U0307 COMBINING DOT ABOVE	
¸	U0327 COMBINING CEDILLA	
˛	U0328 COMBINING OGONEK	
_	U0304 COMBINING MACRON	

---

[14] ⟨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 (U0146, '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⟩

[15] Shapes may vary according to fonts and styles

[16] Position and name of the corresponding diacritical mark in ISO/IEC 10646–1:1993. If possible, combining diacritical marks are referenced. 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

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

[18] ⟨NOTE⟩ This order must be chosen to guarantee compatibility with the Unicode rules for canonical equivalence even though it is at variance with Greek practice. This argument is irrelevant for MES–2, but the problem must be faced for MES–3⟨/NOTE⟩

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

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

[21] 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 chosen to enable the unification of diacritical marks accross scripts without modifying the resulting sequence of strings. The Unicode concept of canonical equivalence implies the necessity to do so.

A.8.1.2 Multiple diacritical marks

Table 2 – Multiple diacritical marks

Shape	Corresponding diacritical mark in ISO/IEC 10646[22]
	Multiple diacritical marks for the Greek script
ʾ	U1FCE PSILI AND OXIA
	— PSILI AND OXIA AND PROSGEGRAMMENI
	— PSILI AND OXIA AND YPOGEGRAMMENI
ʿ	U1FCD PSILI AND VARIA
	— PSILI AND VARIA AND PROSGEGRAMMENI
[23] ~	— PSILI AND VARIA AND YPOGEGRAMMENI
	U1FCF PSILI AND PERISPOMENI
	— PSILI AND PERISPOMENI AND PROSGEGRAMMENI
	— PSILI AND PERISPOMENI AND YPOGEGRAMMENI
	— PSILI AND PROSGEGRAMMENI
	— PSILI AND YPOGEGRAMMENI
ʿ	U1FDE DASIA AND OXIA
	— DASIA AND OXIA AND PROSGEGRAMMENI
	— DASIA AND OXIA AND YPOGEGRAMMENI
ʿ	U1FDD DASIA AND VARIA
	— DASIA AND VARIA AND PROSGEGRAMMENI
	— DASIA AND VARIA AND YPOGEGRAMMENI
~	U1FDF 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
∴	U1FED DIALYTIKA AND VARIA
~	U1FEE DIALYTIKA AND OXIA
∴	U1FC1 DIALYTIKA AND PERISPOMENI
	U0385 DIALYTIKA AND TONOS
́	Multiple diacritical marks for the Latin script
̈	— RING ABOVE AND ACUTE
̈	— DIAERESIS AND MACRON

---

[22] 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

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

A.8.2 Equivalent forms

Table 3 – Equivalent forms

Shape	Position and name in ISO/IEC 10646	Equivalent
æ	U00E6 LATIN SMALL LETTER AE	ae
Æ	U00C6 LATIN CAPITAL LETTER AE	Ae
ǽ	U01FD LATIN SMALL LETTER AE WITH ACUTE	áe
Ǽ	U01FC LATIN CAPITAL LETTER AE WITH ACUTE	Áe
dz	U01F3 LATIN SMALL LETTER DZ	dz
DZ	U01F1 LATIN CAPITAL LETTER DZ	DZ
dž	U01C6 LATIN SMALL LETTER DZ WITH CARON	dž
DŽ	U01C4 LATIN CAPITAL LETTER DZ WITH CARON	DŽ
fi	UFB01 LATIN SMALL LIGATURE FI	fi
fl	UFB02 LATIN SMALL LIGATURE FL	fl
ĳ	U0133 LATIN SMALL LIGATURE IJ	ij
Ĳ	U0132 LATIN CAPITAL LIGATURE IJ	IJ
lj	U01C9 LATIN SMALL LETTER LJ	lj
LJ	U01C7 LATIN CAPITAL LETTER LJ	LJ
nj	U01CC LATIN SMALL LETTER NJ	nj
NJ	U01CA LATIN CAPITAL LETTER NJ	NJ
œ	U0153 LATIN SMALL LIGATURE OE	oe
Œ	U0152 LATIN CAPITAL LIGATURE OE	Oe

A.8.3 Modified letters

Table 4 – Modified letters

Shape	Position and name of modified letter in ISO/IEC 10646	Equiv. POL[24]
đ	U0111 LATIN SMALL LETTER D WITH STROKE	d
Đ	U0110 LATIN CAPITAL LETTER D WITH STROKE	D
ð	U00F0 LATIN SMALL LETTER ETH	d
Ð	U00D0 LATIN CAPITAL LETTER ETH	D
ǥ	U01E5 LATIN SMALL LETTER G WITH STROKE	g
Ǥ	U01E4 LATIN CAPITAL LETTER G WITH STROKE	G
ħ	U0127 LATIN SMALL LETTER H WITH STROKE	h
Ħ	U0126 LATIN CAPITAL LETTER H WITH STROKE	H
ı	U0131 LATIN SMALL LETTER DOTLESS I	i
İ	U0130 LATIN CAPITAL LETTER I WITH DOT ABOVE	I
	U0192 LATIN SMALL LETTER F WITH HOOK	f
	U0138 LATIN SMALL LETTER KRA	k
ł	U0142 LATIN SMALL LETTER L WITH STROKE	l
Ł	U0141 LATIN CAPITAL LETTER L WITH STROKE	L
l·	U0140 LATIN SMALL LETTER L WITH MIDDLE DOT	l
L·	U013F LATIN CAPITAL LETTER L WITH MIDDLE DOT	L
ⁿ	U207F SUPERSCRIPT LATIN SMALL LETTER N	n
’n	U0149 LATIN SMALL LETTER N PRECEDED BY APOSTROPHE	n
	U014B LATIN SMALL LETTER ENG	ng
	U014A LATIN CAPITAL LETTER ENG	NG
ø	U00F8 LATIN SMALL LETTER O WITH STROKE	o
Ø	U00D8 LATIN CAPITAL LETTER O WITH STROKE	O
	U027C LATIN SMALL LETTER R WITH LONG LEG	r
ſ	U017F LATIN SMALL LETTER LONG S	s
ß	U00DF LATIN SMALL LETTER SHARP S	ss
ŧ	U0167 LATIN SMALL LETTER T WITH STROKE	t
Ŧ	U0166 LATIN CAPITAL LETTER T WITH STROKE	T
ʒ	U0292 LATIN SMALL LETTER EZH	z
Ʒ	U01B7 LATIN CAPITAL LETTER EZH	Z
ǯ	U01EF LATIN SMALL LETTER EZH WITH CARON	z
Ǯ	U01EE LATIN CAPITAL LETTER EZH WITH CARON	Z

---

[24] Equivalant on Primary Ordering Level

Annex B (informative): Word-by-word ordering

B.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.

B.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.

B.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

B.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 (A.3.3.1) spacing characters precede digits and letters. The space character is then removed from the table of special characters. The other ordering rules remain unchanged.

Annex C (informative): Ordering by position and by style[25]

C.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 typographic features, e. g. italic typeface, that serves to indicate some property of the word.

C.2 Recommended rules

This European Standard recommends that, if the implementer 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.) small letter on baseline
2.) capital letter on baseline
3.) small letter above baseline
4.) capital letter above baseline
5.) small letter below baseline
6.) capital 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 D (informative): Mixed-script ordering with one predominant script (»encyclopedia mode«)

D.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.

---

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

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

D.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 suggests that strings be transliterated according to the relevant ISO standards. Some of these stardards are currently being revised by ISO TC46.